

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

 Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

 (TM)

 Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Mparc_hntp n.a. - n.a. Smith-Waterman search, using a protein database
 which has been backtranslated into n.a. using IUPAC symbols

Run on: Fri Apr 16 14:29:02 1999; Maspar time 95.81 Seconds

Tabular output not generated. 944.391 Million cell updates/sec

Title: >US-09-011-797-1
 Description: (1-932) from US09011797.seq
 Perfect Score: 4660 1 CTCCTCCAGCGCTTCAGCAG.....AAAAAAAAAAAAAGAAATTC 932
 N.A. Sequence: GAGAGGCGCAGCAAGTCGTC.....TTTTTTTTTTTCCTTAAG
 Comp:

Scoring table: TABLE bkttranslated2
 Gap 30

Nmatch STD : Dbase 0; Query 0

Searched: 131922 seqs, 48541980 bases x 2

Post-processing: Minimum Match 08
 Listing first 45 summaries

Database:

a-geneeq32
 1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
 8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
 14:part14 15:part15 16:part16 17:part17 18:part18
 19:part19 20:part20 21:part21 22:part22 23:part23
 24:part24 25:part25 26:part26 27:part27 28:part28
 29:part29

Statistics: Mean 57.111; Variance 273.971; scale 0.208

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	199	4.3	17 24	W25163	Human opioid receptor	4.50e-03
2	196	4.2	17 25	W25163	Rat orphanin FQ recep	7.17e-03
3	196	4.2	17 24	W25162	Human opioid receptor	7.17e-03
4	181	3.9	17 25	W25220	Rat orphanin FQ recep	7.06e-02
5	175	3.8	1719 16	R2100	Human R1Z allele D283	1.73e-01
6	170	3.6	562 14	R70491	Leucocytotoxin protoso	3.61e-01
7	170	3.6	712 25	W30749	Rat vnt21 gene produc	3.61e-01
8	166	3.6	4987 3	R10834	Rianodin receptor.	6.48e-01
9	161	3.5	392 27	W30091	FIN13 serine/chretonin	1.33e+00
10	160	3.4	58 4	PA0312	PIG brain thalamus po	1.54e+00
11	160	3.4	249 17	R85900	Human pp32 protein.	1.54e+00
12	157	3.4	412 19	W03626	Human thyrotropin GR	2.36e+00
13	155	3.3	182 17	R85902	Mouse pp32 partial se	3.13e+00
14	155	3.3	182 4	R21740	pp35-related protein.	3.13e+00

15	155	3.3	707 16	R79912	Human nucleolin.	3.13e+00
16	156	3.3	2986 8	R44432	eryA region polypepti	2.72e+00
17	151	3.2	542 28	W36490	Human TUP1 protein.	5.48e+00
18	148	3.2	875 28	W34987	Bankia gouldi endoglu	8.31e+00
19	146	3.1	118 2	R12349	Toxoplasma gondii F2	1.09e+01
20	143	3.1	162 6	R30641	pcENP-B-1 protein jac	1.65e+01
21	145	3.1	289 8	R52633	Guinea pig pR-30, 30	1.25e+01
22	143	3.1	594 7	R34936	CENP-B.	1.65e+01
23	146	3.1	725 27	W39165	Human RHAMM protein.	1.09e+01
24	144	3.1	729 1	R05457	KEX1 amino acid sequ	1.44e+01
25	141	3.0	19 24	W00941	Acidmax acidic exten	2.16e+01
26	142	3.0	98 24	W00951	bHx Max with acidic	1.89e+01
27	140	3.0	226 9	R46485	Rat GAP-43.	2.47e+01
28	140	3.0	226 6	R25332	Sequence encoded by r	2.47e+01
29	140	3.0	226 6	R05907	Rat neuronal growth r	2.47e+01
30	140	3.0	226 6	R30015	Rat GAP-43.	2.47e+01
31	140	3.0	226 9	R46496	Rat GAP-43.	2.47e+01
32	140	3.0	227 1	R05908	Mouse neuronal growth	2.47e+01
33	140	3.0	283 27	W37482	Mouse liver cancer-or	2.47e+01
34	138	3.0	414 20	W10529	Saccharomyces cerevis	3.23e+01
35	140	3.0	905 26	W31186	Human p160 polypeptid	2.47e+01
36	140	3.0	1132 17	R97866	Chicken leucocytotoxin	2.47e+01
37	140	3.0	1135 26	W31185	Human p160 polypeptid	2.47e+01
38	136	2.9	27 23	W11629	Anti-apoptotic protei	4.21e+01
39	135	2.9	27 17	R94328	Anti-thermogenic pep	4.80e+01
40	135	2.9	252 1	P94369	Fusion protein congl.	4.80e+01
41	135	2.9	288 29	W41574	Human cardiac tropoin	4.80e+01
42	135	2.9	352 23	W22437	Frog neurogenic diffe	4.80e+01
43	135	2.9	360 19	W03627	Human follicle stimu	4.80e+01
44	135	2.9	484 29	W43910	Heterodera glycines s	4.80e+01
45	135	2.9	1427 2	R10534	Human 160KD mediator	4.80e+01

ALIGNMENTS

RESULT 1
 ID W25163 standard: peptide: 17 AA.
 AC W25163;
 DT 08-DEC-1997 (first entry)
 DE Human opioid receptor-like 1 receptor ligand, nociceptin, fragment.
 KW Opioid receptor-like 1; ORL1; ligand; pronociceptive; stress;
 KM hyperalgesia; locomotor activity disease; anxiety; neuroendocrine;
 KW memory; attention; sensory perception; learning; homeostasis;
 KW hypoalgesia; nociceptin; endorphin; dynorphin A.
 OS Synthetic.
 PN MO9707208-A1.
 PD 27-FEB-1997.
 PF 14-AUG-1986; BE0087.
 PR 15-AUG-1995; US-002366.
 PA (ULBR) UNIV LIBRE BRUXELLES.
 PI Meunier J, Mollereau C, Parmentier M, Vassart G;
 DR WPI: 97-165292/15.
 DR N-PSDB: T79888.
 PT Novel ligand for the opioid receptor-like receptor, nociceptin -
 PT has pro-nociceptive properties, useful for treating or preventing
 PT diseases related to e.g. stress, hyperalgesia, locomotor activity,
 PT etc.
 PS claim 7: Page 27; 48pp; English.
 CC W25162-W25164 are fragments of a ligand to human opioid receptor-like 1
 CC (ORL1) receptor, designated nociceptin, which resembles the endorphin
 CC dynorphin A. Inhibitors of nociceptin or its DNA, e.g. antisense
 CC sequences and antibodies can be used as new types of drugs in the
 CC control of various behaviours or functions. The inhibitors can be used
 CC to prevent or treat a disease related to hyperalgesia, neuroendocrine
 CC secretion, stress, locomotor activity, anxiety, instinctive behaviour,
 CC learning and memory, homeostasis, hyperalgesia, hypoalgesia and/or
 CC sensory perception.
 SQ Sequence 17 AA;
 Query Match 4.3%; Score 199; DB 24; Length 17;
 Best Local Similarity 49.0%; Pred. No. 4.50e-03;
 Matches 25; Conservative 18; Mismatches 8; Indels 0; Gaps 0;

Db 1 f s e f m r q y l v l s m q s s q
 Dc 1 f s e f m r q y l v l s m q s s q
 OY 421 TTGAGTGTATGAGGACGACCTGCTGAGCAAGCAACCA 471
 Ct F S E F M R Q Y L V L S M Q S S Q

RESULT 2
 ID W25219 standard; peptide; 17 AA.
 AC W25219;
 DT 12-JAN-1998 (first entry)
 DE Rat orphanin FQ receptor-binding peptide inhibitor.
 KW Orphanin FQ receptor; binding; locomotor disease; diagnosis;
 KW treatment; opioid inhibitor; opiate induced hypothermia; drug design;
 OS Synthetic.
 PN W09707/212-A1.
 PD 27-FEB-1997.
 PR 12-AUG-1996; U13305.
 PR 03-NOV-1995; US-553058.
 PA (UYOR-) UNIV OREGON HEALTH SCI.
 PI Bunzow JR, Civeill O, Grandy DK, Grisel JE, Mogil JS;
 DR MPI: 97-165296/15.
 PT Mammalian orphanin FQ receptor inhibitor - used in the diagnosis and
 PS treatment of locomotor disease.
 CC Claim 12; Page 48; 68pp; English.
 CC W25219 and W25220 are peptide inhibitors of the rat orphanin FQ (ORF)
 CC receptor.
 CC Highly specific peptides that bind the ORF receptor were identified as
 CC ORF receptor inhibitors. The peptides can be used to antagonise a
 CC physiological effect of an opioid in an animal. The peptides antagonise a
 CC opiate induced hypothermia and morphine induced analgesia in animals.
 CC They may also be used in the diagnosis and treatment of locomotor
 CC disease. The peptides may also be used in the design of a methadone
 CC specific opioid receptor (MSOR), in drug design and for the isolation
 CC of endogenous receptors for anti-opioid agonists and antagonists found
 CC in plasma, serum, lymph, cerebrospinal fluid etc.
 SO Sequence 17 AA.

Query Match
 Best Local Similarity 4.2%; Score 196; DB 25; Length 17;
 Matches 29; Conservative 11; Mismatches 11; Indels 0; Gaps 0;
 Db 1 f g g f t g a r k s a r k l a n g
 Dc 1 f g g f t g a r k s a r k l a n g
 OY 364 TTGGGGCTTCACTGGGCGGAGTCAAGTGGCCACCAAG 414
 Ct F G G F T G A R K S A R K L A N Q

RESULT 3
 ID W25162 standard; peptide; 17 AA.
 AC W25162;
 DT 08-DEC-1997 (first entry)
 DE Human opioid receptor-like 1 receptor ligand, nociceptin, fragment.
 KW Opioid receptor-like 1; ORL1; ligand; pronociceptive; stress;
 KW memory; attention; sensory perception; anxiety; neuroendocrine;
 KW hypoaesthesia; nociceptin; endorphin; dynorphin A.
 OS Synthetic.
 PN W09707/208-A1.
 PD 27-FEB-1997.
 PR 14-AUG-1996; BE0087.
 PR 15-AUG-1995; US-002368.
 PA (ULBR-) UNIV LIBRE BRUXELLES.
 PI Meunier J, Molleireau C, Parmentier M, Vassart G;
 DR MPI: 97-165292/15.
 PT Novel 1996.
 CC has pro-nociceptive properties, useful for treating or preventing
 PT diseases related to e.g. stress, hyperalgesia, locomotor activity,

PT etc.
 PS Claim 5; Page 27; 48pp; English.
 CC W25162-W25164 are fragments of a ligand to human opioid receptor-like 1
 CC (ORL1) receptor, designated nociceptin, which resembles the endorphin
 CC dynorphin A. Inhibitors of nociceptin or its DNA, e.g. antisense
 CC sequences and antibodies of nociceptin or its DNA, e.g. antisense
 CC control of various behaviours or functions. The inhibitors can be used
 CC to prevent or treat a disease related to hyperalgesia, neuroendocrine
 CC secretion, stress, locomotor activity, anxiety, instinctive behaviour,
 CC learning and memory, homeostasis, hyperalgesia, hypoaesthesia and/or
 CC sensory perception.
 SO Sequence 17 AA.

Query Match
 Best Local Similarity 4.2%; Score 196; DB 24; Length 17;
 Matches 29; Conservative 11; Mismatches 11; Indels 0; Gaps 0;
 Db 1 f g g f t g a r k s a r k l a n g
 Dc 1 f g g f t g a r k s a r k l a n g
 OY 364 TTGGGGCTTCACTGGGCGGAGTCAAGTGGCCACCAAG 414
 Ct F G G F T G A R K S A R K L A N Q

RESULT 4
 ID W25220 standard; peptide; 17 AA.
 AC W25220;
 DT 12-JAN-1998 (first entry)
 DE Rat orphanin FQ receptor-binding peptide inhibitor.
 KW Orphanin FQ receptor; binding; locomotor disease; diagnosis;
 KW treatment; opioid inhibitor; opiate induced hypothermia; drug design;
 OS Synthetic.
 PN W09707/212-A1.
 PD 27-FEB-1997.
 PR 12-AUG-1996; U13305.
 PR 03-NOV-1995; US-553058.
 PA (UYOR-) UNIV OREGON HEALTH SCI.
 PI Bunzow JR, Civeill O, Grandy DK, Grisel JE, Mogil JS;
 DR MPI: 97-165296/15.
 PT Mammalian orphanin FQ receptor inhibitor - used in the diagnosis and
 PS treatment of locomotor disease.
 CC Claim 12; Page 48; 68pp; English.
 CC W25219 and W25220 are peptide inhibitors of the rat orphanin FQ (ORF)
 CC receptor.
 CC Highly specific peptides that bind the ORF receptor were identified as
 CC ORF receptor inhibitors. The peptides can be used to antagonise a
 CC physiological effect of an opioid in an animal. The peptides antagonise a
 CC opiate induced hypothermia and morphine induced analgesia in animals.
 CC They may also be used in the diagnosis and treatment of locomotor
 CC disease. The peptides may also be used in the design of a methadone
 CC specific opioid receptor (MSOR), in drug design and for the isolation
 CC of endogenous receptors for anti-opioid agonists and antagonists found
 CC in plasma, serum, lymph, cerebrospinal fluid etc.
 SO Sequence 17 AA.

Query Match
 Best Local Similarity 3.9%; Score 181; DB 25; Length 17;
 Matches 29; Conservative 10; Mismatches 12; Indels 0; Gaps 0;
 Db 1 f g g f t g a r k s a r k l a n g
 Dc 1 f g g f t g a r k s a r k l a n g
 OY 364 TTGGGGCTTCACTGGGCGGAGTCAAGTGGCCACCAAG 414
 Ct F G G F T G A R K S A R K L A N Q

RESULT 5
 ID R92100 standard; Protein; 1719 AA.
 AC R92100;
 DT 19-MAY-1996 (first entry)

Wed Apr 28 09:08:27 1999

US-09-011-797-1.rag

Page 7

```
Dbb      e e e e e e  
Dc       argargargargargarg   810  
Qy      :|||:|||:|||:  
Qt    338 TGGACGAGAACGCCTCAGAAA   360  
        E Q K Q L Q K
```

Search completed: Fri Apr 16 14:34:06 1999
Job time : 304 secs.

THIS PAGE BLANK (USPTO)

 WISE (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPsrch_n n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Apr 27 11:27:17 1999; Maspar time 148.56 Seconds

Tabular output not generated. 853.547 Million cell updates/sec

Title: >US-09-011-797-1
 Description: (1-932) from US09011797.seq
 Perfect Score: 932
 N.A. Sequence: 1 CTCCTCCACCGCTGTCACGAG.....AAAAAAGGAAATTC 932
 Comp: GAGAGGTGCGACAGATGCTC.....TTTTTTTTTTCTTAAAG

Scoring table: TABLE default

Match STD : Dbase 0; Query 0

Searched: 188442 segs; 68026449 bases x 2

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: n-geneseq32
 1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
 8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
 14:part14 15:part15 16:part16 17:part17 18:part18
 19:part19 20:part20 21:part21 22:part22 23:part23
 24:part24 25:part25 26:part26 27:part27 28:part28
 29:part29 30:part30 31:part31 32:part32 33:part33
 34:part34 35:part35 36:part36 37:part37 38:part38
 39:part39 40:part40

Statistics: Mean 9.182; Variance 6.819; scale 1.347

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	932	100.0	932	34	T79888 DNA encoding human op	0.00e+00
2	114	12.2	186	10	076890 Human genome fragment	6.28e-45
3	43	4.6	91	9	051746 Oligonucleotide probe	3.67e-08
4	43	4.6	204	1	N81164 Base substituted E.co	1.09e-07
5	42	4.5	204	1	N81164 Base substituted E.co	1.09e-07
6	41	4.4	91	9	051746 Oligonucleotide probe	7.80e-06
7	38	4.1	114	12	070468 Genetic DNA sequence	2.23e-05
8	37	4.0	114	12	070470 Genetic DNA sequence	6.33e-05
9	36	3.9	114	12	070467 Genetic DNA sequence	6.33e-05
10	36	3.9	114	12	070465 Genetic DNA sequence	6.33e-05
11	36	3.9	114	12	070472 Genetic DNA sequence	6.33e-05
12	36	3.9	114	12	070468 Genetic DNA sequence	6.33e-05
13	36	3.9	114	12	070467 Genetic DNA sequence	6.33e-05

Result ID	Score	Query Match	Length	DB ID	Description	Pred. No.
14	35	3.8	114	12	070469 Generic DNA sequence	1.78e-04
15	34	3.6	114	12	070465 Generic DNA sequence	4.97e-04
16	34	3.6	114	12	070470 Generic DNA sequence	1.37e-03
17	33	3.5	114	12	070466 Generic DNA sequence	1.37e-03
18	33	3.5	114	12	070465 Generic DNA sequence	3.76e-03
19	32	3.4	114	12	070473 Generic DNA sequence	1.02e-02
20	31	3.3	114	12	070472 Generic DNA sequence	1.02e-02
21	31	3.3	114	12	070471 Generic DNA sequence	2.73e-02
22	30	3.2	114	12	070471 Generic DNA sequence	2.73e-02
23	30	3.2	178	32	T76405 HSV-1 gB and surround	7.22e-02
24	29	3.1	3871	2	N71302 Generic DNA sequence	1.89e-01
25	28	3.0	114	12	070471 Generic DNA sequence	1.89e-01
26	28	3.0	114	12	070473 Generic DNA sequence	1.89e-01
27	28	3.0	178	32	T76405 Human endothelin-1 an	1.89e-01
28	28	3.0	264	32	T76445 Substance P receptor	1.89e-01
29	27	2.9	39	7	051787 Mixed oligonucleotide	4.87e-01
30	27	2.9	82	21	T13610 DC43 T5AR library gen	4.87e-01
31	27	2.9	162	32	T76307 Human RANP5 antisens	4.87e-01
32	27	2.9	172	32	T76363 Human interleukin 8 a	4.87e-01
33	27	2.9	487	2	Q11686 Sequence encoding hum	4.87e-01
34	27	2.9	524	2	Q12088 C-terminal sequence e	1.24e+00
35	26	2.8	608	2	Q12089 Human Ews gene clone	1.24e+00
36	26	2.8	2371	9	Q50643 Human MDNCR antisense	3.10e+00
37	26	2.8	3736	26	T45937 Murine Cp-10.	3.10e+00
38	25	2.7	168	32	T76270 T. ovis vaccine candi	3.10e+00
39	25	2.7	433	4	Q23343 OV-16 antigen.	3.10e+00
40	25	2.7	600	12	Q72739 Partial Factor VII CD	3.10e+00
41	25	2.7	822	3	Q14832 M. hyponeumoniae 2.3k	3.10e+00
42	25	2.7	2177	2	V090695 Human EAA3d excitator	3.10e+00
43	25	2.7	2334	39	V091158 Human EAA3d excitator	3.10e+00
44	25	2.7	3217	12	O81158 Human EAA3d excitator	3.10e+00
45	25	2.7	3385	12	O81156 Human EAA3d excitator	3.10e+00

ALIGNMENTS

RESULT 1
 ID T79888 standard; DNA: 932 BP.
 AC T79888;
 DT 08-DEC-1997 (first entry)
 DE DNA encoding human opiod receptor-like 1 receptor ligand.
 KW Opioid receptor-like 1; ORL1; ligand; pronociceptive; stress;
 KW hyperalgesia; locomotor activity disease; anxiety; neuroendocrine;
 KW memory; attention; sensory perception; learning; homeostasis;
 KW hypoaesthesia; nociceptin; endorphin; dynorphin A; ss.
 OS Synthetic.
 FH Key
 FT cds Location/Qualifiers
 FT 1..507
 FT /*tag= a
 FT /note= "no start codon"
 PN WO9707208-A1.
 PD 27-FEB-1997.
 PF 14-AUG-1996; BE0087.
 PR 15-AUG-1995; US-002368
 PA (ULBR) UNITV LIBRE BRUXELLES.
 PI Meunier J, Mollereau C, Parmentier M, Vassart G;
 DR WPI; 97-165292/15.
 DR P-PSDB; W25162, W25163, W25164.
 PT Novel ligand for the opiod receptor-like receptor, nociceptin -
 has pro-nociceptive properties, useful for treating or preventing
 diseases related to e.g. stress, hyperalgesia, locomotor activity,
 etc.
 PT Claim 1; Page 7; 48pp; English.
 PS T79888 encodes a ligand to human opiod receptor-like 1 (ORL1) receptor
 designated nociceptin, which resembles the endorphin dynorphin A.
 CC Inhibitors of nociceptin or its DNA, e.g. antisense sequences and
 antibodies can be used as new types of drugs in the control of various
 behaviours or functions. The inhibitors can be used to prevent or treat
 a disease related to hyperalgesia, neuroendocrine secretion, stress,
 locomotor activity, anxiety, instinctive behaviour, learning and memory,
 homeostasis, hyperalgesia, hypoaesthesia and/or sensory perception.
 CC Sequence 932 BP; 224 A; 273 C; 238 G; 197 T;

Query Match 100.0%; Score 932; DB 34; Length 932;
 Best Local Similarity 100.0%; Pred. No. 0.00e+00;
 Matches 932; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

1 cctccagcgtgttcacagcgtgtcccgagcgtcctcactgtccagagagaggtccac 60
1 CTCTCAGAGGTTTACAGAGCTGTCCGAGACTGCTTACTGCTCCAGAGAGGTTCCAC 60
61 ccgagtcgggagcgttcacactgaagctgtgcatcctccagtgtagaagaaggtcttc 120
61 CCGGCTCGGGGAGCTTCACTGAGAGCTGTGATCTTCACTGTGAAGAGAGGTTCTTC 120
61 ccccgccctctgtgactctttgacacaaagcagtgagcagtgactctgagcagtcagc 180
121 ccccgccctctgtgactctttgacacaaagcagtgagcagtgactctgagcagtcagc 180
121 ccccgccctctgtgactctttgacacaaagcagtgagcagtgactctgagcagtcagc 180
181 cctgctgacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 240
181 CCGTGTGATCCAGAGCTCAGCTCCGCTGCTTTACAGTGTGAAGAGCCTCGAGATGACG 240
241 caccggaagagaatgcccgtgtcagagtggtgtgcaagcccgagcagcagcagcagc 300
241 CACCTGAAGAGATGCCGCTGTGAGAGTGTGTGAAGCCCGAGAGCGAGAGCTGTAG 300
301 gcagatgcagagcctgtcagatgagcagatgagcagtgagtgagcagcagcagcagc 360
301 GCAGATGCAGAGCCTGTGAGATGAGCCATGAGTGTGAGAGAGAGAGAGAGAGAGAG 360
361 aggtttgggagcctcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 420
361 AGGTTGGGGGCTTCACTGAGGCGCGGAGAGTCAAGCCGGAAGTTGGCCAAACAGAG 420
421 ttcagtgagttatgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 480
421 TTCACTGAGTTATGAGGAGCACTACCTGCTCTGAGCAGTGAAGCAACCGCGCGC 480
481 actctgcacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 540
481 ACTCTGCACAGAGTGTGATGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
541 tgcacccatgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 600
541 TGCAACCATGAGATCAGATCAGATGAGCCCGCTACAGCATGTGTCAACCAAGACTGCA 600
601 ggcgggagtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 660
601 GCGCGGAGTACAGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 660
661 gcatgtcaccacacacacacacacacacacacacacacacacacacacacacacacac 720
661 GCATGTCTCACCAACATCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 720
721 acatttaagcccaatctctgtcaccatcctcgtcagcagcagcagcagcagcagcagc 780
721 ACAATTAAATGCCCCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 780
781 gaacatgaatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 840
781 GAACATGAATCAGAGACTGGGGTTTCCACACACAGCCATACATGCTTTGTAAGAGAG 840
841 ctgtctcttttgactgtgttgaaacacacttctccattaaacttactactagcagaa 900
841 CTGTCTCTTTTGAAGTGTGTTGTAACCACTTCTCCATTAACTTCTACTAGAGAAA 900
901 TGGTTAATAAAAAAAAAAAAAAAAAAGGAAATTC 932
901 TGGTTAATAAAAAAAAAAAAAAAAAAGGAAATTC 932

```

RESULT 2
 ID Q76890 standard; DNA; 186 BP.
 AC Q76890;
 DT 23-SEP-1994 (first entry)

DE Human genome fragment.
 KW Brain; placenta; bone marrow; genetic analysis; gene mapping;
 KW detection; homology; human; adrenal tissue; ds.
 OS Homo sapiens.
 PN M09401346-A.
 PF 20-JAN-1994.
 PF 13-JUL-1993; G01467.
 PR 13-JUL-1992; GB-014857.
 PA (MEDL-) MEDICAL RES COUNCIL.
 PI Gross J, Hatfield KM, Howells D, Kelly M, Shaw D;
 PI Slibson DR, Starkey M;
 DR WPI: 94-035056/04.
 PT New nucleic acid fragment encoding gene products - can be used
 PT for genetic analysis and mapping
 PS Claim 1; Page 256; 616pp; English.
 CC Human nucleic acid fragments, isolated from brain adrenal tissue,
 CC the placenta or bone marrow, comprise any of: (A) a sequence
 CC selected from (Q76401-Q77613), (B) an allelic variation of a
 CC sequence as described in (A), or (C) a sequence complementary
 CC to (A) or (B).
 SQ Sequence 186 BP; 36 A; 63 C; 56 G; 31 T;

Query Match 12.2%; Score 114; DB 10; Length 186;
 Best Local Similarity 81.3%; Pred. No. 6.28e-45;
 Matches 148; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

```

4 aacctgagtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 63
79 MACCTAAGCTGTGCACTCTCTCAGTGTGAAGAGAGTTCCTCCGCGCTCTCTGAGCT 138
64 ccatgacacaaagtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 123
139 CTCTGACCAACAGCCATGCGCAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 198
124 gtggcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 183
199 ACCTCCGCTGCTCTTACAGTGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 258
184 cg 185
259 CG 260

```

RESULT 3
 ID Q51746 standard; cDNA; 91 BP.
 AC Q51746;
 DT 31-MAY-1994 (first entry)
 DE Oligonucleotide probe MK14-A
 KW Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
 KW ss.
 OS Synthetic.
 PN EP-571911-A.
 PD 01-DEC-1993.
 PF 24-MAY-1993; 108325.
 PR 26-MAY-1992; US-889651.
 PA (BECT) BECTON DICKINSON CO.
 PI Shank DD, Spears PA;
 DR WPI: 93-378844/48.
 PT New oligo:nucleotide probes specific for Mycobacteria - used for
 PT detection and amplification of Mycobacteria nucleic acid in
 PT samples
 PS Claim 3; Page 14; 23pp; English.
 CC Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
 CC (Q51735). It hybridized to all spp. of mycobacteria tested, but
 CC cross reacted to a few non-mycobacterial spp. The probe may
 CC be useful as an initial screen for mycobacterial infection.
 CC See also Q51735-45 and Q51747-59.
 SQ Sequence 91 BP; 5 A; 17 C; 15 G; 4 T;

Query Match 4.6%; Score 43; DB 9; Length 91;
 Best Local Similarity 9.1%; Pred. No. 3.67e-08;
 Matches 5; Conservative 44; Mismatches 6; Indels 0; Gaps 0;


```
FT primer_bind 187..204
FT /*tag= b
```

CP 241 GCTGCATCTCCGAGGCTTTCGACTGGTAAGAGCAGCGGACGTGAGCTTGA 189

CC histidine residues incorporated into variant sequences. TSARs are
CC concatenated heterofunctional proteins or peptides, comprising at least
CC two functional regions - a binding domain with affinity for a ligand and
CC a second effector peptide portion that is chemically or biologically
CC active. They may further comprise a linker peptide between the 2 domains.
CC The TSARs or comps. comprising a TSAR binding domain can be used in
CC vivo to deliver a chemically or biologically active moiety, eg. metal
CC ion, radioisotope, peptide, toxin or enzyme, to the specific target or
CC on the cell. They can also replace the function of macromolecules, eg.
CC monoclonal or polyclonal antibodies and therefore circumvent the need
CC for complex methods of hybridoma formation or in vivo antibody
CC production. The TSARs are easily characterised and have designed actively
CC allowing direct and rapid detection in a screening process.
SQ Sequence 114 BP; 6 A; 12 C; 0 G; 0 T;

Query Match 3.9%; Score 36; DB 12; Length 114;
Best Local Similarity 11.6%; Pred. No. 6.33e-05;
Matches 13; Conservative 29; Mismatches 70; Indels 0; Gaps 0;

Db 3 bncanbnnbnnbnnbnnbnnbnnbnnbnnbncanbncanbnnbnnnnnnnn 62
Cp 370 CCCCAACCTTTCTGCAAGCTCTCTCCACCTCATCGCCTCATCTCGACAGGCT 311
Db 63 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnb 114
Cp 310 CTGCATCTGCTCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 259

RESULT 12
ID 070468 standard; DNA: 114 BP.
AC 070468;
DT 05-APR-1995 (first entry)
DE Generic DNA sequence to generate a random TSAR peptide library.
KW TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW effector domain; concatenated heterofunctional protein; linker;
KW direct; rapid; detection; screening; treatment; generic; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT misc_feature 55..60 /*tag= a
FT /note= "this sequence represents 'Z'; Z can be a
FT sequence of 6, 9 or 12 nucleotides (see
FT comments)"

PN W09418318-A.
PD 18-AUG-1994.
PE 01-FEB-1994; U00977.
PR 01-FEB-1993; US-013416.
PR 30-DEC-1993; US-176500.
PR 31-JAN-1994; US-189331.
PA (UYNC-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Kay BK;
DR WPI: 94-279739/34.
DR P-PSDB; R65154.
PT Identifying proteins or peptide(s) which bind a ligand - by
PT screening a recombinant vector library expressing fusion proteins
PS comprising a binding domain and an effector domain
PS Disclosure; Page 35; 253pp; English.
CC 070468 is a generic DNA sequence used to generate random TSAR (Totally
CC Synthetic Affinity Reagents) peptides. This generic formula can also be
CC represented as follows: X(NNB)11(TGC)(NNB)16(TGC)(NNB)10Y. X
CC and Y are flanking restriction sites (X is not the same as Y) that are
CC not specified further. Other generic sequences are shown in 070466-68.
CC Other specific peptides generated by these generic sequences are shown in
CC R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC comprising at least two functional regions - a binding domain with
CC affinity for a ligand and a second effector peptide portion that is
CC chemically or biologically active. They may further comprise a linker
CC peptide between the 2 domains. The oligonucleotides are also designed so
CC that the expressed peptide contains 2 or 4 cysteine residues positioned
CC in, or flanking, the unpredicted or variant residues. These residues
CC confer some degree of conformational rigidity to the peptides. The TSARs
CC or comps. comprising a TSAR binding domain can be used in vivo to
CC deliver a chemically or biologically active moiety, eg. metal ion,

CC radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC cell. They can also replace the function of macromolecules, eg.
CC monoclonal or polyclonal antibodies and therefore circumvent the need
CC for complex methods of hybridoma formation or in vivo antibody
CC production. The TSARs are easily characterised and have designed actively
CC allowing direct and rapid detection in a screening process.
SQ Sequence 114 BP; 0 A; 2 C; 2 G; 2 T;

Query Match 3.9%; Score 36; DB 12; Length 114;
Best Local Similarity 3.6%; Pred. No. 6.33e-05;
Matches 4; Conservative 34; Mismatches 74; Indels 0; Gaps 0;

Db 3 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnb 62
Cp 361 TTTTCTGCAAGCTCTCTCTCCACCTCATCGCCTCATCTCGACAGGCTGCTG 302
Db 63 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnb 114
Cp 301 CCTCAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 250

RESULT 13
ID 070467 standard; DNA: 114 BP.
AC 070467;
DT 05-APR-1995 (first entry)
DE Generic DNA sequence to generate a random TSAR peptide library.
KW TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW effector domain; concatenated heterofunctional protein; linker;
KW direct; rapid; detection; screening; treatment; generic; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT misc_feature 55..60 /*tag= a
FT /note= "this sequence represents 'Z'; Z can be a
FT sequence of 6, 9 or 12 nucleotides (see
FT comments)"

PN W09418318-A.
PD 18-AUG-1994.
PE 01-FEB-1994; U00977.
PR 01-FEB-1993; US-013416.
PR 30-DEC-1993; US-176500.
PR 31-JAN-1994; US-189331.
PA (UYNC-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Kay BK;
DR WPI: 94-279739/34.
DR P-PSDB; R65153.
PT Identifying proteins or peptide(s) which bind a ligand - by
PT screening a recombinant vector library expressing fusion proteins
PS comprising a binding domain and an effector domain
PS Disclosure; Page 35; 253pp; English.
CC 070467 is a generic DNA sequence used to generate random TSAR (Totally
CC Synthetic Affinity Reagents) peptides. This generic formula can also be
CC represented as follows: X(NNB)16(TGC)(NNB)12(TGC)(NNB)11Y. X
CC and Y are flanking restriction sites (X is not the same as Y) that are
CC not specified further. Other generic sequences are shown in 070466-68.
CC Other specific peptides generated by these generic sequences are shown in
CC R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC comprising at least two functional regions - a binding domain with
CC affinity for a ligand and a second effector peptide portion that is
CC chemically or biologically active. They may further comprise a linker
CC peptide between the 2 domains. The oligonucleotides are also designed so
CC that the expressed peptide contains 2 or 4 cysteine residues positioned
CC in, or flanking, the unpredicted or variant residues. These residues
CC confer some degree of conformational rigidity to the peptides. The TSARs
CC or comps. comprising a TSAR binding domain can be used in vivo to
CC deliver a chemically or biologically active moiety, eg. metal ion,
CC radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC cell. They can also replace the function of macromolecules, eg.
CC monoclonal or polyclonal antibodies and therefore circumvent the need for
CC complex methods of hybridoma formation or in vivo antibody production.
CC The TSARs are easily characterised and have designed actively allowing
CC direct and rapid detection in a screening process.
SQ Sequence 114 BP; 0 A; 2 C; 2 G; 2 T;

[illegible]

Wed Apr 28 09:08:29 1999

US-09-011-797-1.rng

Page 8

Search completed: Tue Apr 27 11:29:51 1999
Job time : 154 secs.

 M I T O C H O N D R I A L
 (TM)

Release 3.1a John F. Collins, BioComputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPearch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Fri Apr 16 14:18:44 1999; MasPar time 2.64 Seconds
 Tabular output not generated. 173.116 Million cell updates/sec

Title: >US-09-011-797-2
 Description: (1-17) from US09011797.pep
 Perfect Score: 116
 Sequence: 1 FGFTGARKSARKLANQ 17

Scoring table:
 PAM 150
 Gap 15

Searched: 74019 segs, 26840295 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: swiss-prot35
 1:swissprot

Statistics: Mean 27.528; Variance 32.859; scale 0.838

Prod. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	116	100.0	17	PNOC_PIG	NOCICEPTIN (ORPHANIN F	3.09e-15
2	116	100.0	17	PNOC_HUMAN	NOCICEPTIN PRECURSOR (3.09e-15
3	116	100.0	181	PNOC_RAT	NOCICEPTIN PRECURSOR (3.09e-15
4	116	100.0	181	PNOC_MOUSE	NOCICEPTIN PRECURSOR (3.09e-15
5	67	57.8	205	RS4_MYCN	30S RIBOSOMAL PROTEIN	1.44e-02
6	66	56.9	205	RS4_MYCN	30S RIBOSOMAL PROTEIN	2.41e-02
7	61	52.6	405	RS4_PEA	GLYCERALDEHYDE 3-PHOS	2.94e-01
8	60	51.7	310	PNAD_PIG	PROTEIN N-TERMINAL ASP	4.77e-01
9	60	51.7	310	PNAD_MOUSE	PROTEIN N-TERMINAL ASP	4.77e-01
10	60	51.7	752	CUL1_HUMAN	CULLIN HOMOLOG 1 (CUL	4.77e-01
11	59	50.9	773	LIT13_DROME	LIN-13 HOMOLOG PROTEIN	7.70e-01
12	58	50.0	208	RS4_HELPY	30S RIBOSOMAL PROTEIN	1.24e+00
13	58	50.0	421	NASA_BACSU	NITRATE TRANSPORTER.	1.24e+00
14	58	50.0	998	CBF_HUMAN	CCNA1-BINDING FACTOR (1.24e+00
15	57	49.1	159	YICN_ECOLI	HYPOTHETICAL 18.2 KD P	1.97e+00
16	57	49.1	528	PRPR_ECOLI	PROPRIONATE CATABOLISM	1.97e+00
17	57	49.1	541	PRPR_SALTY	PROPRIONATE CATABOLISM	1.97e+00
18	56	48.3	306	CSK1_SCHPO	PROTEIN KINASE CSK1 (E	4.92e+00
19	55	47.4	264	RS2_HELPY	30S RIBOSOMAL PROTEIN	3.12e+00
20	55	47.4	375	EAT3_CABEL	PUTATIVE SODIUM-DEPEND	4.92e+00
21	54	46.6	374	DNAL_CLOAB	DNAJ PROTEIN.	7.70e+00
22	54	46.6	475	Y294_MYCPN	HYPOTHETICAL PROTEIN M	7.70e+00
23	53	45.7	91	DBH1_RHILE	DNA-BINDING PROTEIN HR	1.20e+01

RESULT ID	1	STANDARD:	PRT:	17 AA.	ALIGNMENTS
AC	PNOC_PIG				
AC	P55791:				
DT	01-NOV-1997 (REL. 35, CREATED)				
DT	01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)				
DT	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)				
DE	NOCICEPTIN (ORPHANIN FO).				
GN	PNOC.				
OS	SUS SCROFA (PIG).				
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;				
OC	EUTHERIA; ARTIODACTYLA.				
RN	[1]				
RP	SEQUENCE.				
RC	TISSUE-HYPOTHALAMUS;				
RX	MEDLINE: 96055113;				
RA	REINSCHEID R.K., NOTACKER H.-P., BOURSON A., ARDATI A.,				
RA	HENNINGSSEN R.A., BUNZOW J.R., GRANDY D.K., LANGEN H., MONSMA F.J. JR.,				
RA	GIVELLI O.;				
RL	SCIENCE 270:792-794(1995).				
CC	- FUNCTION: NOCICEPTIN IS THE LIGAND OF THE OPIOID RECEPTOR-LIKE				
CC	RECEPTOR (OPR1). IT MAY ACT AS A TRANSMITTER IN THE BRAIN BY				
CC	MODULATING NOCICEPTIVE AND LOCOMOTOR BEHAVIOR. MAY BE INVOLVED IN				
CC	NEURONAL DIFFERENTIATION AND DEVELOPMENT. WHEN ADMINISTERED				
CC	INTRACEREBROVENTRICULARLY INTO MICE, THIS PEPTIDE INDUCES				
CC	HYPALGESIA AND DECREASES LOCOMOTOR ACTIVITY.				
CC	- TISSUE SPECIFICITY: PREDOMINANTLY IN THE BRAIN AND SPINAL CORD.				
CC	- SIMILARITY: BELONGS TO THE OPIOIDS NEUROPEPTIDES PRECURSORS				
CC	CC FAMILY.				
DR	PROSITE: PS01252; OPIOIDS PRECURSOR; PARTIAL.				
KW	OPIOID PEPTIDE; NEUROTRANSMITTER; NEUROPEPTIDE.				
SQ	SEQUENCE 17 AA: 1809 MW; 09F05DDE CRC32;				
Query Match					100.0%; Score 116; DB 1; Length 17;
Best local Similarity 100.0%; Pred. No. 3.09e-15;					
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
Db	1 FGFTGARKSARKLANQ 17				
QY	1 FGFTGARKSARKLANQ 17				
RESULT 2					
ID	PNOC_HUMAN	STANDARD:	PRT:	176 AA.	
AC	013519:				
DT	01-NOV-1997 (REL. 35, CREATED)				
DT	01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)				


```

DE DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE NOCICEPTIN PRECURSOR (ORPHANIN FQ) (PPNOC).
GN PNOIC OR OFQ.
OC HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 96323281.
RA MOLLEREAU C., SIMONS M.-J., SOULARUE P., LINERS F., VASSART G.,
RA MEDTIER J.-C., PARMENTIER M.;
RA PROC. NATL. ACAD. SCI. U.S.A. 93:8666-8670(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 97312464.
RA LAZAROVI S., PICONE D., TANCREDI T., GUERRINI R., SPADACCINI R.,
RA BIOCHEMA. BIOPHYS. RES. COMMUN. 233:640-643(1997).
RN [3]
RP STRUCTURE BY NMR OF NOCICEPTIN.
RX MEDLINE: 97312464.
RA SALVADORI S., REGOLI D., TEMOSSI P.A.;
RA BIOCHEMA. BIOPHYS. RES. COMMUN. 233:640-643(1997).
RN [4]
RP FUNCTION: NOCICEPTIN IS THE LIGAND OF THE OPIOID RECEPTOR-LIKE
RX RECEPTOR (OPRL). IT MAY ACT AS A TRANSMITTER IN THE BRAIN BY
MODULATING NOCICEPTIVE AND LOCOMOTOR BEHAVIOR. MAY BE INVOLVED
IN NEURONAL DIFFERENTIATION AND DEVELOPMENT (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: SECRETED INTO THE ENDOPLASMIC RETICULUM.
CC -1- TISSUE SPECIFICITY: PREDOMINANTLY IN THE BRAIN AND SPINAL CORD.
CC -1- PTM: SPECIFIC ENZYMATIC CLEAVAGES AT PAIRED BASIC RESIDUES
PROBABLY YIELD OTHER ACTIVE PEPTIDES BESIDES NOCICEPTIN.
CC -1- PTM: THE N-TERMINAL CONTAINS 6 CONSERVED CYSTEINES THOUGHT TO
BE INVOLVED IN DISULFIDE BONDING AND/OR PROCESSING.
CC -1- SIMILARITY: BELONGS TO THE OPIOIDS NEUROPEPTIDES PRECURSORS
FAMILY.
DR EMBL: X87370; E244777; -
DR EMBL: X87367; E244785; -
DR EMBL: X87368; E244785; JOINED.
DR EMBL: 048263; G1185010; -
DR MIM: 601459; -
DR PROSITE: PS01252; OPIOIDS PRECURSOR; 1.
KW OPIOID PEPTIDE; NEUROTRANSMITTER; NEUROPEPTIDE;
KM CLEAVAGE ON PAIR OF BASIC RESIDUES; SIGNAL.
FT SIGNAL 1 19
FT PEPTIDE 98 127 NEUROPEPTIDE 1 (PROBABLE).
FT PEPTIDE 130 146 NOCICEPTIN (ORPHANIN FQ).
FT PEPTIDE 149 165 NEUROPEPTIDE 2 (PROBABLE).
SQ SEQUENCE 176 AA; 20295 MW; 8055843A CRC32;

Query Match 100.0%; Score 116; DB 1; Length 176;
Best Local Similarity 100.0%; Pred. No. 3, 09e-15;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0
OY 1 FGGFTGARRSARKLANQ 17

RESULT 3
ID PNOIC_RAT STANDARD; PRT; 181 AA.
AC Q62923; 064162;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE NOCICEPTIN PRECURSOR (ORPHANIN FQ) (PPNOC) (ORL1 RECEPTOR AGONIST).
GN PNOIC.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-SPRAGUE-DAWLEY;

```

```

RA NOTACER H.P., REINSCHIED R.K., MANSOUR A., HENNINGSEN R.A.,
RA MONSIEU F.J., WATSON S.J., CIVELLI O.,
RL SUBMITTED (FEB-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 96323281.
RA MOLIEREAU C., SIMONS M.-J., SOULARUE P., LINERS F., VASSART G.,
RA MEUNIER J.-C., PARMENTIER M.,
RL PROC. NATL. ACAD. SCI. U.S.A. 93:8666-8670(1996).
[3]
RP SEQUENCE OF 13-181 FROM N.A., AND SEQUENCE OF 135-151.
RC TISSUE-BRAIN;
RX MEDLINE: 96013745.
RA MEUNIER J.-C., MOLIEREAU C., TOLL L., SUAUDEAU C., MOISAND C.,
RA ALVIERIE P., BATOIR J.-L., GUILLEMOT J.-C., FERRARA P.,
RA MONSARAT B., MAZAGUI H., VASSART G., PARMENTIER M., COSTENTIN J.,
RL NATURE 377:532-535(1995).
CC -1 FUNCTION: NOCICEPTIN IS THE LIGAND OF THE OPIOID RECEPTOR-LIKE
CC RECEPTOR (OPRL). IT MAY ACT AS A TRANSMITTER IN THE BRAIN BY
CC MODULATING NOCICEPTIN AND LOCOMOTOR BEHAVIOR. MAY BE INVOLVED
CC IN NEURONAL DIFFERENTIATION AND DEVELOPMENT.
CC -1 SUBCELLULAR LOCATION: SECRETED INTO THE ENDOPLASMIC RETICULUM.
CC -1 TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN THE SPINAL CORD
CC AND BRAIN, BEING MORE ABUNDANT IN THE HIPOTHALAMUS AND STRIATUM.
CC ALSO FOUND IN SMALL AMOUNTS IN OVARY.
CC -1 PTM: SPECIFIC ENYMATIC CLEAVAGES AT PAIRED BASIC RESIDUES
CC PROBABLY YIELD OTHER ACTIVE PEPTIDES BESIDES NOCICEPTIN.
CC -1 PTM: THE N-TERMINAL CONTAINS 6 CONSERVED CYSTEINES THOUGHT TO
CC BE INVOLVED IN DISULFIDE BONDING AND/OR PROCESSING.
CC -1 SIMILARITY: BELONGS TO THE OPIOIDS NEUROPEPTIDES PRECURSORS
CC FAMILY.
DR EMBL: U48262; GI185012; -.
DR EMBL: X97375; E244795; -.
DR EMBL: S79730; G1172239; -.
KW PROSITE: PS01252: OPIOIDS_PRECURSOR, 1.
KM OPIOID PEPTIDE; NEUROTRANSMITTER; NEUROPEPTIDE;
KM CLEAVAGE ON PAIR OF BASIC RESIDUES; SIGNAL; REPEAT.
FT SIGNAL 1
FT PEPTIDE 98 132 NEUROPEPTIDE 1 (PROBABLE).
FT PEPTIDE 135 151 NOCICEPTIN (ORPHANIN FQ).
FT PEPTIDE 154 170 NEUROPEPTIDE 2 (PROBABLE).
FT DOMAIN 109 120 2 X 6 AA TANDEN REPEATS OF D-A-E-P-X-A.
FT REPEAT 109 114
FT REPEAT 115 120 1.
SQ SEQUENCE 181 AA; 20478 MW; 1640366F CRC32;

Query Match 100.0%; Score 116; Length 181;
Best Local Similarity 100.0%; Pred. No. 3, Dbl-15;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

Db 135 FGGFTGARKSARKLANQ 151
QY 1 FGGFTGARKSARKLANQ 17
|||||
|||||

RESULT 4
ID PNOC_MOUSE STANDARD: PRT: 187 AA.
AC Q64387; O61105; O61938;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE NOCICEPTIN PRECURSOR (ORPHANIN FQ) (PNOC) (N23K / N27K).
OS PNOC OR NPNC1.
OS MUS MUSCULUS (MUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
[1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE: 96216718.
RA HOUTANI T., NISHI M., TAKESHIMA H., NUKADA T., SUGIMOTO T.,
RA BLUMCHEN. BIOPHYS. RES. COMMUN. 215:714-719(1996).
[2]

```


RP SEQUENCE FROM N.A.
 RX MEDLINE: 96106851.
 RA SATO Y., MARUYAMA K., SAITO T.C., KAWASHIMA S.;
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 217:539-545(1995).
 RN (3)
 RP SEQUENCE FROM N.A.
 RC STRAIN-129/SV;
 RX MEDLINE: 96323281.
 RA MOLLEREAU C., SIMONS M.-J., SOULARUE P., LINERS F., VASSART G.,
 RL MENIER J.-C., PARMENTIER M.;
 RA PROC. NATL. ACAD. SCI. U.S.A. 93:8666-8670(1996).
 RN (4)
 RP SEQUENCE FROM N.A. (LONG FORM).
 RX MEDLINE: 96279082.
 RA SATO Y., MARUYAMA K., KAWANO H., HAGINO-YAMAGISHI K.,
 RA KAWAMURA K., SAITO T.C., KAWASHIMA S.;
 RL J. BIOL. CHEM. 271:15615-15622(1996).
 RN (5)
 RP TISSUE-BRAIN: 79-187 FROM N.A.
 RC MEDLINE: 96207555.
 RA PAN Y.-X., XU J., PASTERNAK G.M.;
 RL BIOCHEM. J. 315:11-13(1996).
 CC -1- FUNCTION: NOCICEPTIN IS THE LIGAND OF THE OPIOID RECEPTOR-LIKE
 RECEPTOR (OPRL). IT MAY ACT AS A TRANSMITTER IN THE BRAIN BY
 MODULATING NOCICEPTIVE AND LOCOMOTOR BEHAVIOR. MAY BE INVOLVED
 IN NEURONAL DIFFERENTIATION AND DEVELOPMENT. WHEN ADMINISTERED
 INTRACEREBROVENTRICULARLY, NOCICEPTIN INDUCES HYPERALGESIA AND
 DECREASES LOCOMOTOR ACTIVITY.
 CC -1- SUBCELLULAR LOCATION: SECRETED INTO THE ENDOPLASMIC RETICULUM.
 CC -1- TISSUE SPECIFICITY: BRAIN AND SPINAL CORD. LOW LEVELS IN
 KIDNEY AND SPLEEN.
 CC -1- DEVELOPMENTAL STAGE: IN EMBRYONIC BRAIN, FIRST DETECTED AT DAY 14
 AND IN POSTNATAL BRAIN, LEVELS INCREASE IN DAY 1 AND DAY 18.
 CC -1- LEVELS DECREASE SIGNIFICANTLY IN ADULTS.
 CC -1- PPM: SPECIFIC ENZYMAIC CLEAVAGES AT PAIRED BASIC RESIDUES
 PROBABLY YIELD OTHER ACTIVE PEPTIDES BESIDES NOCICEPTIN.
 CC -1- PPM: THE N-TERMINAL CONTAINS 6 CONSERVED CISTEINES THOUGHT TO
 BE INVOLVED IN DISULFIDE BONDING AND/OR PROCESSING.
 CC -1- ALTERNATIVE PRODUCTS: A LONGER FORM, N27K IS PROBABLY PRODUCED
 BY ALTERNATIVE SPLICING.
 CC -1- SIMILARITY: BELONGS TO THE OPIOIDS NEUROPEPTIDES PRECURSORS
 FAMILY.
 DR EMBL: D82866; D1012281; -;
 DR EMBL: D50056; D1009395; -;
 DR EMBL: X97373; E244874; -;
 DR EMBL: X97371; E244782; -;
 DR EMBL: X97372; E244782; JOINED.
 DR EMBL: D50055; G1304162; -;
 DR EMBL: U44027; G1335970; -;
 DR MGD: MGI:105308; NPNC1.
 DR PROSITE: PS01252; OPIOIDS; PRECURSOR; 1.
 KW OPIOID PEPTIDE; NEUROTRANSMITTER; NEUROPEPTIDE; ALTERNATIVE SPLICING;
 KW CLEAVAGE ON PAIR OF BASIC RESIDUES; SIGNAL; REPEAT.
 FT SIGNAL 1 19
 FT PEPTIDE 98 138
 FT PEPTIDE 141 157
 FT PEPTIDE 160 176
 FT DOMAIN 109 126
 FT REPEAT 109 114
 FT REPEAT 115 120
 FT REPEAT 121 126
 FT REPEAT 121 126
 FT VARSPLIC 186 187
 FT CONFLICT 80 80
 FT SEQUENCE 187 AA; 20884 MM; F810ABF0 CRC32;
 NV -> IQVPRICAVHSCRCPCGVRIIPSPRH (IN
 LONG FORM; N27K).
 V -> T (IN REF. 5).
 Query Match 100.0%; Score 116; DB 1; Length 187;
 Best Local Similarity 100.0%; Pred. No. 3,09e-15;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 DB 141 FGFTGARKSARKLANO 157
 |||

OY 1 FGFTGARKSARKLANO 17
 RESULT 5
 ID RS4 MYCPN STANDARD; PRT; 205 AA.
 AC P46775;
 DT 01-NOV-1995 (REL. 32, CREATED)
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE 30S RIBOSOMAL PROTEIN S4.
 GN RPSD OR RPS4.
 OS MYCOPLASMA PNEUMONIAE.
 CC PROKARYOTA; TENERICUTES; MOLICUTES; MYCOPLASMA; MYCOPLASMA TALE;
 CC MYCOPLASMA TALE.
 CC (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 29342 / M129;
 RA DIRKSEN I.B., PROFT T., HILBERT H., PLAGENS H., HERRMANN R.,
 RA KRAUSE D.C.;
 RL SUBMITTED (MAR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 29342 / M129;
 RX MEDLINE: 97105885.
 RA HIMELEFELCH R., HILBERT H., PLAGENS H., PIRKL E., LI B.-C.,
 RA HERRMANN R.;
 RL NUCLEIC ACIDS RES. 24:4420-4449(1996).
 CC -1- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 16S RIBOSOMAL RNA
 (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE S4P FAMILY OF RIBOSOMAL PROTEINS.
 DR EMBL: L38897; G639791; -;
 DR EMBL: AE000038; G1674077; -;
 DR PROSITE: PS00632; RIBOSOMAL_S4; 1.
 KW RIBOSOMAL PROTEIN; RNA-BINDING.
 SW SEQUENCE 205 AA; 23774 MM; 165319A8 CRC32;
 Query Match 57.8%; Score 67; DB 1; Length 205;
 Best Local Similarity 53.3%; Pred. No. 1.44e-02;
 Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
 DB 105 GFAPTRRSAROLVNH 119
 OY 3 GFCTGARKSARKLANO 17
 RESULT 6
 ID RS4 MYCPN STANDARD; PRT; 205 AA.
 AC P47553;
 DT 01-FEB-1996 (REL. 33, CREATED)
 DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
 DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
 DE 30S RIBOSOMAL PROTEIN S4.
 GN RPSD OR RPS4 OR MG311.
 OS MYCOPLASMA GENTRALIUM.
 CC PROKARYOTA; TENERICUTES; MOLICUTES; MYCOPLASMA; MYCOPLASMA TALE;
 CC MYCOPLASMA TALE.
 CC (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 33530 / G-37;
 RX MEDLINE: 96026346.
 RA FRASER C.M., GOCAYNE J.D., WHITE O., ADAMS M.D., CLAYTON R.A.,
 RA FLEISCHMANN R.D., BULT C.J., KERLAVAGE A.R., SUTTON G., KELLEY J.M.,
 RA FRITZMAN J.L., WEIDMAN J.F., SMALL K.V., SANDUSKY M., FUHRMANN J.L.,
 RA NGUYEN D.T., UTERBACK T.R., SADEK D.M., PHILLIPS C.A., MERRICK J.M.,
 RA TOMB J.-F., DOUGHERTY B.A., BOTT K.F., HU P.-C., LUCIER T.S.,
 RA PETERSON S.N., SMITH H.O., HUTCHISON C.A. III, VENTER J.C.;
 RL SCIENCE 270:397-403(1995).
 CC -1- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 16S RIBOSOMAL RNA
 (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE S4P FAMILY OF RIBOSOMAL PROTEINS.
 DR EMBL: U39712; G1046011; -;
 DR TIGR: MG311; -;
 DR PROSITE: PS00632; RIBOSOMAL_S4; 1.

KW RIBOSOMAL PROTEIN; RNA-BINDING.
SQ SEQUENCE 205 AA; 23947 MW; 02D16821 CRC32;

Query Match 56.9%; Score 66; DB 1; Length 205;
Best Local Similarity 53.3%; Pred. No. 2 41e-02;
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 105 GFAPTRKSRKRVNH 119
||:|||||:|
QY 3 GFTGAKRSARKLANQ 17

RESULT 7
ID G3PA_PEA STANDARD; PRT; 405 AA.
AC P12858;
DT 01-OCT-1989 (REL. 12, CREATED)
DT 01-OCT-1992 (REL. 22, LAST SEQUENCE UPDATE)
DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
DE GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE A PRECURSOR, CHLOROPLAST
DE (EC 1.2.1.12).
GN GAPA OR GPAL.
OS PISUM SATIVUM (GARDEN PEA).
OC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE; FABALES;
OC FABACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. ROSAKRONE; TISSUE-SEEDLING;
RX MEDLINE; 91062390.
RA LIAUD M.-F., ZHANG D.-X., CEREF R.;
RL PROC. NATL. ACAD. SCI. U.S.A. 87:8918-8922(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. ROSAKRONE; TISSUE-SEEDLING;
RX MEDLINE; 93357435.
RA BRINKMANN H., CEREF R., SALOMON M., SOLL J.;
RL PLANT MOL. BIOL. 13:81-94(1989).
CC -1- CATALYTIC ACTIVITY: D-GLYCERALDEHYDE 3-PHOSPHATE + ORTHOPHOSPHATE
+ NAD(+) -> 1,3-DIPHOSPHATEGLYCERATE + NADH.
CC -1- PATHWAY: CALVIN CYCLE.
CC -1- SUBUNIT: TETRAMER OF EITHER FOUR A CHAINS (GAPDH 2) OR TWO A AND
TWO B CHAINS (GAPDH 1).
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST.
CC -1- PLANTS CONTAIN THREE FORMS OF GAPDH: A CYTOSOLIC FORM WHICH
PARTICIPATES IN GLYCOLYSIS AND TWO CHLOROPLAST FORMS WHICH
PARTICIPATES IN PHOTOSYNTHESIS. THESE THREE FORMS ARE ENCODED BY
DISTINCT GENES.
EMBL: X52148; G12159; -;
DR EMBL; X15190; G20729; -;
DR PIR; S14243; DEPMNA.
DR HSSP; P00362; 1GDI.
DR PROSITE; PS00071; GAPDH. 1.
KW CALVIN CYCLE; OXIDOREDUCTASE; NAD; TRANSIT PEPTIDE; CHLOROPLAST;
KM MULTIGENE FAMILY.
FT TRANSIT 1 68
FT CHAIN 69 405
FT FT 68
FT BINDING 222 222
FT ACT_SITE 249 249
FT CONFLICT 194 194
SQ SEQUENCE 405 AA; 43338 MW; A9F53CE6 CRC32;

Query Match 52.6%; Score 61; DB 1; Length 405;
Best Local Similarity 46.7%; Pred. No. 2 94e-01;
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Db 20 FSEFGLRNSRHLP 34
|:|:|:|:|:|:|
QY 1 FCGTGAARKSARKLA 15

RESULT 8
ID PNAD_PIG STANDARD; PRT; 310 AA.
AC Q28955;

DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE PROTEIN N-TERMINAL ASPARAGINE AMIDOHYDROLASE (EC 3.5.1.-) (PROTEIN
DE NH2-TERMINAL ASPARAGINE DEAMIDASE) (NTN-AMIDASE) (PNAD).
GN NTN1.
OS SUS SCROFA (PIG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; ARTIODACTYLA.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-LIVER.
RX MEDLINE; 95113832.
RA STEWART A.E., ARFIN S.M., BRADSHAW R.A.;
RL J. BIOL. CHEM. 270:25-28(1995).
RN [2]
RP CHARACTERIZATION.
RC TISSUE-LIVER.
RX MEDLINE; 94375451.
RA STEWART A.E., ARFIN S.M., BRADSHAW R.A.;
RL J. BIOL. CHEM. 269:23509-23517(1994).

-1- FUNCTION: SIDE CHAIN DEAMIDATION OF N-TERMINAL ASPARAGINE RESIDUES
TO ASPARTIC ACID. REQUIRED FOR THE UBQUITIN-DEPENDENT TURNOVER OF
INTRACELLULAR PROTEINS THAT INITIATE WITH MET-ASN. THESE PROTEINS
ARE ACETYLATED ON THE RETAINED INITIATOR METHIONINE AND CAN
SUBSEQUENTLY BE MODIFIED BY THE REMOVAL OF N-ACETYL METHIONINE BY
ACYLAMINOACID HYDROLASE. CONVERSION OF THE RESULTING NH2-TERMINAL
ASPARAGINE TO ASPARTIC ACID BY PNAD RENDERS THE PROTEIN
SUSCEPTIBLE TO ARGINYLTATION, POLYUBQUITINYLTATION AND DEGRADATION
AS SPECIFIED BY THE N-END RULE.
CC -1- SUBUNIT: MONOMER.
DR EMBL; U17062; G595950; -;
KW HYDROLASE.
SQ SEQUENCE 310 AA; 34760 MW; 3110E5ED CRC32;

Query Match 51.7%; Score 60; DB 1; Length 310;
Best Local Similarity 43.8%; Pred. No. 4 77e-01;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Db 127 GGFTRKRSARKLANQ 142
||:|:|:|:|:|
QY 2 GGFTRKRSARKLANQ 17

RESULT 9
ID PNAD_MOUSE STANDARD; PRT; 310 AA.
AC Q64311;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE PROTEIN N-TERMINAL ASPARAGINE AMIDOHYDROLASE (EC 3.5.1.-) (PROTEIN
DE NH2-TERMINAL ASPARAGINE DEAMIDASE) (NTN-AMIDASE) (PNAD).
GN NTN1.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; ROSENTIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C129;
RX MEDLINE; 97067079.
RA GRIGORYEV S., STEWART A.E., KWON Y.T., ARFIN S.M., BRADSHAW R.A.,
RA JERKINS N.A., COPELAND N.G., VARSHAVSKY A.;
RL J. BIOL. CHEM. 271:28521-28532(1996).
RN [2]
RP FUNCTION: SIDE CHAIN DEAMIDATION OF N-TERMINAL ASPARAGINE RESIDUES
TO ASPARTIC ACID. REQUIRED FOR THE UBQUITIN-DEPENDENT TURNOVER OF
INTRACELLULAR PROTEINS THAT INITIATE WITH MET-ASN. THESE PROTEINS
ARE ACETYLATED ON THE RETAINED INITIATOR METHIONINE AND CAN
SUBSEQUENTLY BE MODIFIED BY THE REMOVAL OF N-ACETYL METHIONINE BY
ACYLAMINOACID HYDROLASE. CONVERSION OF THE RESULTING NH2-TERMINAL
ASPARAGINE TO ASPARTIC ACID BY PNAD RENDERS THE PROTEIN
SUSCEPTIBLE TO ARGINYLTATION, POLYUBQUITINYLTATION AND DEGRADATION
AS SPECIFIED BY THE N-END RULE.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).

DR EMBL: U57692; G1373367; -
 DR EMBL: U57691; G1373365; -
 DR MGI: 108471; NTANTL.
 KW HYDROLASE.
 SQ SEQUENCE 310 AA; 34595 MW; 739412CE CRC32;

Query Match 51.7%; Score 60; DB 1; Length 310;
 Best Local Similarity 43.8%; Pred. No. 4,77e-01;
 Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Db 127 GGFSDRQLSOKLTHQ 142
 |||:|::|::|:
 QY 2 GGFCAKRSARKLANQ 17

RESULT 10
 ID CUL1_HUMAN STANDARD: PRT: 752 AA.
 AC 013616;
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE CULLIN HOMOLOG 1 (CUL-1).
 GN CUL1.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 OC EUTHERIA; PRIMATES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 96279828.
 RA KIREOS E.T., LANDER L.E., WING J.P., HE W.W., HEDGECOCK E.M.;
 RL CELL 85:839-839(1996).
 CC -1- SIMILARITY: BELONGS TO THE CULLIN FAMILY.
 DR EMBL: U58087; G1381142; -
 DR PROSITE: PS01256; CULLIN_1; 1.
 DR PROSITE: PS00069; CULLIN_2; 1.
 SQ SEQUENCE 752 AA; 87387 MW; 6CEB00D8 CRC32;

Query Match 51.7%; Score 60; DB 1; Length 752;
 Best Local Similarity 46.7%; Pred. No. 4,77e-01;
 Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Db 542 FTAFASRHSGRKLT 556
 |::|::|::|:
 QY 1 FGFTGAKRSARKLANQ 15

RESULT 11
 ID L119_DROME STANDARD: PRT: 773 AA.
 AC 024311;
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE LIN-19 HOMOLOG PROTEIN.
 GN LIN19.
 OS DROSOPHILA MELANOGASTER (FRUIT FLY).
 OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-OREGON-R; TISSUE-BRAIN;
 RA FILIPPOV V.A., FILIPPOVA M.A., SEHNAL F.;
 RL SUBMITTED (JAN-1996) TO EMBL/GENBANK/DBS DATA BANKS.
 CC -1- SIMILARITY: BELONGS TO THE CULLIN FAMILY.
 DR EMBL: L41642; G1146334; -
 DR FLYBASE: FB990015509; lin19.
 DR PROSITE: PS01256; CULLIN_1; 1.
 DR PROSITE: PS00069; CULLIN_2; 1.
 SQ SEQUENCE 773 AA; 89411 MW; 8851D23F CRC32;

Query Match 50.9%; Score 59; DB 1; Length 773;
 Best Local Similarity 57.1%; Pred. No. 7,70e-01;
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Db 565 FNEFYAHSGRKL 578

QY 1 FGFTGAKRSARKLANQ 14
 |::|::|::|:

RESULT 12
 ID RS4_HELPY STANDARD: PRT: 208 AA.
 AC P56011;
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE 30S RIBOSOMAL PROTEIN S4.
 GN RPSD OR HP1294
 OS HELICOBACTER PYLORI (CAMPYLOBACTER PYLORI).
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA;
 OC AEROBIC; MOTILE; HELICAL AND/OR VIBRIOID.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-26695;
 RX MEDLINE; 97394467.
 RA TOMB J.-F., WHITE O., KERLAVAGE A.R., CLAYTON R.A., SUTTON G.G.,
 RA FLEISCHMANN R.D., KETCHUM K.A., KLENK H.-P., GILL S., DOUGHERTY B.A.,
 RA NELSON K., QUACKENBUSH J., ZHOU L., KIRKNESS E.F., PETERSON S.,
 RA LOFTUS B., RICHARDSON D., DODSON R., KHARAK H.G., GLIDER A.,
 RA MCKENNEY K., FITZGERALD L.M., LEE N., ADAMS M.D., HICKEY E.K.,
 RA BERG D.E., GOCARNE J.D., UTTERBACK T.R., PETERSON J.D., KELLEY J.M.,
 RA COTTON M.D., WEIDMAN J.M., FUJII C., BOWMAN C., WATNEY L., WALLIN E.,
 RA HAYES W.S., BORODOVSKY M., KARP P.D., SMITH H.O., FRASER C.M.,
 RA VENTER J.C.;
 RL NATURE 388:539-547(1997).
 CC -1- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 16S RIBOSOMAL RNA
 (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE S4P FAMILY OF RIBOSOMAL PROTEINS.
 DR EMBL: AE000633; G2314460; -
 DR TIGR: HP1294; -
 DR PROSITE: PS00632; RIBOSOMAL_S4; 1.
 KW RIBOSOMAL PROTEIN; RNA-BINDING.
 SQ SEQUENCE 208 AA; 23964 MW; EE21B30C CRC32;

Query Match 50.0%; Score 58; DB 1; Length 208;
 Best Local Similarity 46.7%; Pred. No. 1,24e+00;
 Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Db 108 GFATRSSARQLVTH 122
 ||::|::|::|:
 QY 3 GFTGAKRSARKLANQ 17

RESULT 13
 ID NASC_BACCU STANDARD: PRT: 421 AA.
 AC P42432;
 DT 01-NOV-1995 (REL. 32, CREATED)
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE NITRATE TRANSPORTER.
 GN NASC.
 OS BACILLUS SUBTILIS.
 OC PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RX MEDLINE; 95173124.
 RA OGAWA K.-I., AKAGAWA E., YAMANE K., SUN Z.-W., LACELLE M., ZUBER P.,
 RA NAKANO M.M.;
 RL J. BACTERIOL. 177:1409-1413(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RX MEDLINE; 97124189.
 RA YAMANE K., KUNANO M., KURITA K.;
 RL MICROBIOLOGY 142:3047-3056(1996).
 CC -1- FUNCTION: MAY FUNCTION AS A NITRATE TRANSPORTER.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO THE NARX/NASA FAMILY OF TRANSPORTERS.

DR EMBL: D30689; G1437473; -
 DR EMBL: D50453; G1805405; -
 DR EMBL: Z99105; E1182285; -
 DR SUBLIST: EG11093; NASA.
 KW TRANSPORT; TRANSMEMBRANE; NITRATE ASSIMILATION.
 FT TRANSMEM 19 39 POTENTIAL.
 FT TRANSMEM 51 71 POTENTIAL.
 FT TRANSMEM 79 99 POTENTIAL.
 FT TRANSMEM 105 125 POTENTIAL.
 FT TRANSMEM 136 156 POTENTIAL.
 FT TRANSMEM 167 187 POTENTIAL.
 FT TRANSMEM 217 237 POTENTIAL.
 FT TRANSMEM 248 268 POTENTIAL.
 FT TRANSMEM 277 297 POTENTIAL.
 SQ SEQUENCE 421 AA; 46067 MW; D7B4779 CRC32;
 Query Match 50.0%; Score 58; DB 1; Length 421;
 Best Local Similarity 61.5%; Pred. No. 1.24e+00;
 Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
 Db 388 GKGARKARKRM 400
 QY 2 GGFTGARKARKL 14

RESULT 14
 ID CBF_HUMAN STANDARD; PRT; 998 AA.
 AC 003701;
 DT 01-OCT-1996 (REL. 34, CREATED)
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
 DE CCAAT-BINDING FACTOR (CBF).
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 CC EUTHERIA; PRIMATES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 91061780.
 RA LUM L., SULTZMAN L., KAUFMAN R., LINZER D.I.H., WU B.;
 RL MOL. CELL. BIOL. 10:6709-6717(1990).
 CC -1- FUNCTION: STIMULATES TRANSCRIPTION FROM THE HSP70 PROMOTER.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 DR EMBL: M37197; G179969; -
 KW TRANSCRIPTION REGULATION; ACTIVATOR; NUCLEAR PROTEIN.
 SQ SEQUENCE 998 AA; 114071 MW; 6DD27B10 CRC32;

Query Match 50.0%; Score 58; DB 1; Length 998;
 Best Local Similarity 46.7%; Pred. No. 1.24e+00;
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 958 GSFOGPRKKRNND 972
 QY 2 GGFTGARKARKL 16

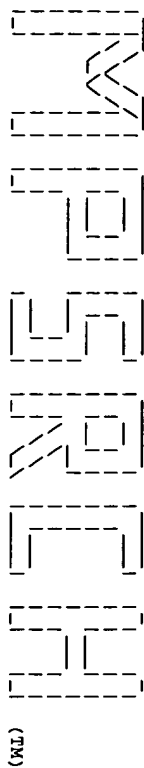
RESULT 15
 ID YICN_ECOLI STANDARD; PRT; 159 AA.
 AC P31439;
 DT 01-JUL-1993 (REL. 26, CREATED)
 DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL 18.2 KD PROTEIN IN NUPA-UHPT INTERGENIC REGION.
 GN YICN
 OS ESCHERICHIA COLI.
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
 CC ENTEROBACTERIACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=K12 / MG1655;
 RX MEDLINE; 93315143.
 RA BURLAND V.D., PLUNKETT G. III, DANIELS D.L., BLATTNER F.R.;
 RL GENOMICS 16:551-561(1993).
 DR EMBL: L10328; G290512; -

DR EMBL: AE000444; G1790096; -
 DR ECOGENE: EG11690; YICN.
 KW HYPOTHETICAL PROTEIN.
 SQ SEQUENCE 159 AA; 18173 MW; F09A1692 CRC32;

Query Match 49.1%; Score 57; DB 1; Length 159;
 Best Local Similarity 46.2%; Pred. No. 1.97e+00;
 Matches 6; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

Db 31 SGARKAIRLSDR 43
 QY 5 TGARKSAKLANQ 17

Search completed: Fri Apr 16 14:18:52 1999
 Job time : 8 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mparch_tpm n.a. - n.a. Smith-Waterman search, using a protein query
which has been backtranslated into n.a. using IUPAC symbols

Run on: Tue Apr 27 10:21:16 1999; Maspar time 137.97 Seconds

Tabular output not generated. 873.545 Million cell updates/sec

Title: >US-09-011-797-3
Description: (1-17) from US09011797.pep
Perfect Score: 215
N.A. Sequence: 1 TTYWSNGARTTATGMCNCA.....TNSNATGCARMSWNSNCAR 51
Comp: AARNSTCYAARFACKNGT.....ANMSNFCAGYMSWNSNGTY

Scoring table: TABLE bkttranslated2

Gap 40

Nmatch STD : Dbase 0; Query 0

Searched: 602357 seqs, 1181590623 bases x 2

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

emb157
1:em_ba 2:em_fun 3:em_hlg 4:em_hum1 5:em_hum2 6:em_in
7:em_om 8:em_or 9:em_ov 10:em_pat 11:em_ph 12:em_pl
13:em_ro 14:em_un 15:em_v1

Database:

genbank110
16:gb_ba1 17:gb_ba2 18:gb_hlg 19:gb_in 20:gb_om 21:gb_ov
22:gb_pat 23:gb_ph 24:gb_pl1 25:gb_pl2 26:gb_pl3
27:gb_pr2 28:gb_pr3 29:gb_ro 30:gb_st 31:gb_sts 32:gb_sy
33:gb_un 34:gb_v1

Statistics: Mean 48.293; Variance 156.627; scale 0.308

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	199	92.6	372	29	MMU44027	Mus musculus orphanin
2	199	92.6	633	29	KMPPNMNA	M.musculus mRNA for pr
3	199	92.6	878	26	HSPPNEX3	H.sapiens gene encodin
4	199	92.6	925	29	S79730	ORL1 receptor agonist
5	199	92.6	972	29	MUSN27K	Mouse N23K mRNA for de
6	199	92.6	972	29	RNU48262	Rattus norvegicus pre-
7	199	92.6	988	29	RMPPNMNA	R.norvegicus mRNA for
8	199	92.6	1015	26	HSPPNMNA	H.sapiens mRNA for pre
9	199	92.6	1047	29	MUSN23K	Mouse N23K mRNA for N2
10	199	92.6	1198	28	HSU48263	Human pre-pro-orphanin
11	199	92.6	1209	20	AB005251	Bos taurus mRNA for no
12	199	92.6	1253	29	KMPPNEX3	M.musculus gene encodi

13	199	92.6	1354	29	MUSNOP	Mouse mRNA for nocicep	3.66e-07
14	124	57.7	294095	18	HS1170K4	Human DNA sequence ***	6.77e+00
15	119	55.3	166983	27	AC002541	Human BAC clone RG043K	1.83e+01
16	118	54.9	32981	19	CE1ZK418	Caenorhabditis elegans	2.22e+01
17	118	54.9	74589	24	AB010068	Arabidopsis thaliana g	2.22e+01
18	117	54.4	156237	18	HS111B22	Human DNA sequence ***	2.70e+01
19	116	54.0	69822	19	AC003121	Drosophila melanogaste	3.28e+01
20	114	53.0	41552	19	CEC17D12	Caenorhabditis elegans	4.81e+01
21	114	53.0	118357	28	HS1142C11	Human DNA sequence fro	5.83e+01
22	113	52.6	1123	24	NT33WOP2	N.tabacum mRNA for 33	5.83e+01
23	113	52.6	1175	24	LE33KOECP	L.esculentum mRNA for	5.83e+01
24	113	52.6	29016	19	CET23B5	Caenorhabditis elegans	5.83e+01
25	113	52.6	95267	18	HS91U24	Human DNA sequence ***	5.83e+01
26	112	52.1	762	19	A1U25289	Aramigus intermedicus c	7.05e+01
27	112	52.1	100635	27	AC003104	Homo sapiens chromosom	7.05e+01
28	112	52.1	141460	18	AC004829	*** SEQUENCING IN PROG	7.05e+01
29	111	51.6	17986	19	CE1ZK75	Caenorhabditis elegans	8.51e+01
30	109	50.7	27576	29	MMGR5505	Mus musculus G protein	1.24e+02
31	109	50.7	30480	19	CE22B3	Caenorhabditis elegans	1.24e+02
32	109	50.7	78025	28	HS780M13	Human DNA sequence fro	1.24e+02
33	109	50.7	86958	27	AC004464	Homo sapiens 12q13.1 p	1.24e+02
34	109	50.7	190778	27	HS142L7	Homo sapiens DNA sequ	1.24e+02
35	108	50.2	180	34	SEV6KPRO	Semliki Forest Virus (1.49e+02
36	108	50.2	3000	29	RNU14908	Rattus norvegicus cyto	1.49e+02
37	108	50.2	3843	34	SEV6K	Semliki Forest Virus R	1.49e+02
38	108	50.2	11517	22	AR001293	Sequence 1 from patent	1.49e+02
39	108	50.2	26154	19	CE33E2	Caenorhabditis elegans	1.49e+02
40	108	50.2	39763	28	AC005853	Homo sapiens clone UMG	1.49e+02
41	108	50.2	39805	27	AC002096	Homo sapiens chromosom	1.49e+02
42	108	50.2	77713	27	HUAC004131	Homo sapiens chromosom	1.49e+02
43	108	50.2	125983	27	HS306D1	Human DNA sequence fro	1.49e+02
44	108	50.2	14474	27	AF051934	Homo sapiens chromosom	1.49e+02
45	108	50.2	348050	27	HS229042	Homo sapiens 959 kb co	1.49e+02

ALIGNMENTS

RESULT	1	LOCUS	MMU44027	372 bp	mRNA	ROD	30-MAY-1996
DEFINITION							
ACCESSION							
NID							
KEYWORDS							
SOURCE							
ORGANISM							
REFERENCE							
AUTHORS							
TITLE							
JOURNAL							
MEDLINE							
REFERENCE							
AUTHORS							
TITLE							
JOURNAL							
FEATURES							
source							
CDS							
misc_feature							

/note="encodes an active 17 amino acid peptide;
FCGFTGARSARLANQ"

BASE COUNT 85 a 105 c 122 g 60 t

ORIGIN

Query Match 92.6%; Score 199; DB 29; Length 372;
Best Local Similarity 49.0%; Pred. No. 3,66e-07;
Matches 25; Conservative 18; Mismatches 8; Indels 0; Gaps 0;

Db 245 TTCAGTAGTTTATGAGCAGTACCTGTCCTGAGCATGACAGTCAAGTCAA 295
1 TTYWSNGARTTYATGMCNARTAYTNGTNTYWSNATGCAWSMNSNCAR 51

RESULT 2
LOCUS MPMRMRNA 633 bp RNA ROD 10-SEP-1996
DEFINITION M.musculus mRNA for prepronociceptin.
ACCESSION X97373
NID 91532039
KEYWORDS prepronociceptin.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 633)
Mollereau,C., Simons,M.J., Soularue,P., Liners,F., Vassart,G., Meunier,J.C. and Parmentier,M.
Structure, tissue distribution, and chromosomal localization of the prepronociceptin gene
Proc. Natl. Acad. Sci. U.S.A. 93 (16), 8666-8670 (1996)

JOURNAL MEDLINE
REFERENCE 2 (bases 1 to 633)
Parmentier,M.
Direct Submission
Submitted (19-APR-1996) M. Parmentier, Universite Libre de Bruxelles, I R I B H N ULB Campus Erasme, 808 Route de Lennik, B-1070 Bruxelles, BELGIUM

AUTHORS TITLE
JOURNAL

FEATURES
SOURCE Location/Qualifiers
1. .633
/organism="Mus musculus"
/strain="129/Sv"
/db_xref="taxon:10090"
/clone_lib="lambda FIXII (Stratagene #946305)"
/clone="MG43"
<1. .150
/number=2
join(<1. .150,151. .>633)
join(25. .150,151. .588)
/codon_start=1
/product="prepronociceptin"
/db_xref="PID:e244874"
/db_xref="PID:91532040"
/translation="MKILFCDYLLLSVSSPPRCINQEKLHAPDSFNKTC
ILQCEKVEPRPLMTVCTKYMASGSGQSLPADPEVSAALTOPKASEQHLKRPVR
SLVOYDAPGADAPGADAPGADAPGADAEVEQKQLOKRFGEFTGARSARLANQKRF
SEFMROYLVLSMOSORRTLHONGNV"
25. .69
70. .585
/product="prepronociceptin"
151. .>633
/number=3

BASE COUNT 138 a 194 c 182 g 119 t

ORIGIN

Query Match 92.6%; Score 199; DB 29; Length 633;
Best Local Similarity 49.0%; Pred. No. 3,66e-07;
Matches 25; Conservative 18; Mismatches 8; Indels 0; Gaps 0;

Db 502 TTCAGTAGTTTATGAGCAGTACCTGTCCTGAGCATGACAGTCAAGTCAA 552
1 TTYWSNGARTTYATGMCNARTAYTNGTNTYWSNATGCAWSMNSNCAR 51

RESULT 3
LOCUS HSPPNEX3 878 bp DNA PRI 10-SEP-1996
DEFINITION H.sapiens gene encoding prepronociceptin, exon 3.
ACCESSION X97368
NID 91531986
KEYWORDS prepronociceptin.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 878)
Mollereau,C., Simons,M.J., Soularue,P., Liners,F., Vassart,G., Meunier,J.C. and Parmentier,M.
Structure, tissue distribution, and chromosomal localization of the prepronociceptin gene
Proc. Natl. Acad. Sci. U.S.A. 93 (16), 8666-8670 (1996)

JOURNAL MEDLINE
REFERENCE 2 (bases 1 to 878)
Parmentier,M.
Direct Submission
Submitted (19-APR-1996) M. Parmentier, Universite Libre de Bruxelles, I R I B H N ULB Campus Erasme, 808 Route de Lennik, B-1070 Bruxelles, BELGIUM

AUTHORS TITLE
JOURNAL

FEATURES
SOURCE Location/Qualifiers
1. .878
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="lambda DASHII (Stratagene #94520)"
/clone="HG91"
/chromosome="8"
/map="p21"
/map="1.9 CR from WI-1172 marker"
73. .524
/number=3
/label=ex3
/usedin=X97367:HSPPN_CDS
/usedin=X97367:HSPPN_mRNA
/product="prepronociceptin"

BASE COUNT 194 a 274 c 231 g 179 t

ORIGIN

Query Match 92.6%; Score 199; DB 26; Length 878;
Best Local Similarity 49.0%; Pred. No. 3,66e-07;
Matches 25; Conservative 18; Mismatches 8; Indels 0; Gaps 0;

Db 391 TTCAGTAGTTTATGAGCAGTACCTGTCCTGAGCATGACAGTCAAGTCAA 441
1 TTYWSNGARTTYATGMCNARTAYTNGTNTYWSNATGCAWSMNSNCAR 51

RESULT 4
LOCUS 579730 925 bp mRNA ROD 30-JAN-1996
DEFINITION ORL1 receptor agonist precursor-endogenous agonist of opioid receptor-like ORL1 receptor [rats, brain, mRNA Partial, 925 nt].
ACCESSION S79730
NID 9117238
KEYWORDS Rattus sp. brain.
SOURCE Rattus sp.
ORGANISM Eukaryota; Eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 925)
Mollereau,C., Toll,L., Snaudeau,C., Moisan,C., Alvinerie,P., Butour,J.L., Guillemot,J.C., Ferrata,P., Monsarrat,B. et al.
Isolation and structure of the endogenous agonist of opioid receptor-like ORL1 receptor [see comments]
Nature 377 (6549), 532-535 (1995)

JOURNAL MEDLINE
REMARK 96013745
Genbank staff at the National Library of Medicine created this entry [NCBI g1bseq 171655] from the original journal article.

OY 1 TTYWSNGARTTATGMCNCARTAYTNGTNTYWSNATGCARWSMNCAR 51

RESULT 7
LOCUS RNPENRNA 988 bp RNA
DEFINITION R.norvegicus mRNA for prepronociceptin.
ACCESSION X97375
NID 91532064
KEYWORDS prepronociceptin.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 988)
AUTHORS Mollereau,C., Simons,M.J., Soularue,P., Liners,F., Vassart,G.,
Meunier,J.C. and Parmentier,M.
TITLE Structure, tissue distribution, and chromosomal localization of the
prepronociceptin gene
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 93 (16), 8666-8670 (1996)
MEDLINE 96323281
REFERENCE 2 (bases 1 to 988)
AUTHORS Parmentier,M.
TITLE Direct Submission
JOURNAL Submitted (19-APR-1996) M. Parmentier, Universite libre de
Bruxelles, I R I B N ULB Campus Erasme, 808 Route de Lennik, B-
1070 Bruxelles, BELGIUM
COMMENT Related sequence 579730.
FEATURES
source location/Qualifiers
1..988
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone_1id="lambda DASHII (Stratagene #943501)"
/clone_1id="lambda ZAPII (brain cDNA)"
/clone="RGE109"
/clone="RB101"
<1..150
/number=2
join(<1..150,151..615,616..974)
25..69
join(25..150,151..570)
/codon_start=1
/product="prepronociceptin"
/db_xref="PID:6244795"
/db_xref="PID:91532065"
/translation="MKILFCVLLSLSSVSSCPEDCLNCOERLHPAPSPYLKIC
ILQCEKVPPLMTCTKAMASDESDSPADPELTSAIYOSRASEMOKRPRVR
SVYQADAEPEADAPVADEADDEVQKQKRRFGFTGARKSARKLANOKRSEFMRO
YVLNMOSSQRRRLHONGNV"
70..567
/product="prepronociceptin"
151..615
/number=3
616..974
/number=4
971
BASE COUNT 231 a 292 c 250 g 215 t
ORIGIN

Query Match 92.6%; Score 199; DB 29; Length 988;
Best Local Similarity 49.0%; Pred. No. 3,666-07;
Matches 25; Conservative 18; Mismatches 8; Indels 0; Gaps 0;

Db 484 TTCAGTGGATTATGAGCAGTCTGCTGAGCATGCGTACAGCCAA 534
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
OY 1 TTYWSNGARTTATGMCNCARTAYTNGTNTYWSNATGCARWSMNCAR 51

RESULT 8
LOCUS HSPENRNA 1015 bp RNA
DEFINITION H.sapiens mRNA for prepronociceptin.
ACCESSION X97370
NID 91531988

KEYWORDS prepronociceptin.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Eukaryota; Eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 1015)
AUTHORS Mollereau,C., Simons,M.J., Soularue,P., Liners,F., Vassart,G.,
Meunier,J.C. and Parmentier,M.
TITLE Structure, tissue distribution, and chromosomal localization of the
prepronociceptin gene
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 93 (16), 8666-8670 (1996)
MEDLINE 96323281
REFERENCE 2 (bases 1 to 1015)
AUTHORS Parmentier,M.
TITLE Direct Submission
JOURNAL Submitted (19-APR-1996) M. Parmentier, Universite libre de
Bruxelles, I R I B N ULB Campus Erasme, 808 Route de Lennik, B-
1070 Bruxelles, BELGIUM
FEATURES
source location/Qualifiers
1..1015
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1id="lambda DASHII (Stratagene #94320)"
/clone="HG91"
/clone="RGE107"
/chromosome="8"
/map="p21"
/map="1.9 CR from WI-1172 marker"
<1..149
/number=2
join(<1..149,150..601,602..>1015)
24..68
join(24..149,150..554)
/codon_start=1
/product="prepronociceptin"
/db_xref="PID:6244777"
/db_xref="PID:91531989"
/translation="MKVILCDLLSLSSVSSCPEDCLNCOERLHPAPSPYLKIC
ILQCEKVPPLMTCTKAMASDESDSPADPELTSAIYOSRASEMOKRPRVR
SVYQADAEPEADAPVADEADDEVQKQKRRFGFTGARKSARKLANOKRSEFMROYLVIS
MOSORRRRLHONGNV"
69..551
/product="prepronociceptin"
150..601
/number=3
602..>1015
/number=4
1015
BASE COUNT 212 a 304 c 264 g 235 t
ORIGIN

Query Match 92.6%; Score 199; DB 26; Length 1015;
Best Local Similarity 49.0%; Pred. No. 3,666-07;
Matches 25; Conservative 18; Mismatches 8; Indels 0; Gaps 0;

Db 468 TTCAGTGGATTATGAGCAGTCTGCTGAGCATGCGTACAGCCAG 518
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
OY 1 TTYWSNGARTTATGMCNCARTAYTNGTNTYWSNATGCARWSMNCAR 51

RESULT 9
LOCUS MUSEN23K 1047 bp RNA
DEFINITION Mouse N23K mRNA for N23K protein, complete cds.
ACCESSION D50056
NID 91857023
KEYWORDS N23K protein; N23K.
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Eukaryotes; Metazoa; Chordata;
Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae;
Murinae; Mus.
REFERENCE 1 (bases 1 to 1047)
AUTHORS Salto,Y., Maruyama,K., Saido,T.C. and Kawashima,S.


```

TITLE
N23K, a gene transiently up-regulated during neural
differentiation, encodes a precursor protein for a newly identified
neuropeptide nociceptin
JOURNAL
Biochem. Biophys. Res. Commun. 217 (2), 539-545 (1995)
MEDLINE
9610681
REFERENCE
2 (bases 1 to 1047)
AUTHORS
Saito, Y.
TITLE
Direct Submission
JOURNAL
Submitted (03-APR-1995) to the DDBJ/EMBL/Genbank databases. Yumiko
Saito, Tokyo Metropolitan Institute for Medical Science, molecular
biology; Honkomagome 3-18-22, Bunkyo-ku, Tokyo 113, Japan
(E-mail:saito@rinsoken.or.jp, Tel:03-3823-2101(ex.5248),
Fax:03-5685-6609)
COMMENT
On Mar 1, 1997 this sequence version replaced gi:1217905.
Sequence updated (06-Mar-1996) by:Yumiko Saito.
FEATURES
source
Location/Qualifiers
1..1047
/db_xref="taxon:10090"
/cell_line="NS20T"
/cell_type="neuroblastoma"
57..620
/gene="N23K"
57..620
/gene="N23K"
/note="neuropeptide precursor"
/codon_start=1
/product="N23K protein"
/db_xref="PID:dl009395"
/db_xref="PID:g1217906"
/translation="MKILFCVLLSLSSVFSFCPRDCLTCQKRLPAPDSFLKTC
ILCEEKVFRLPMTVCTKVASGSGSLSPDPLVSNALQPRASEMQLKRPVR
SLVQVRADEPADAEPGADAEPADEAEQOKLQRFGGFTGARSARKLANQKFF
SEFMROYLVPSMOSSORRLHONGNV"
BASE COUNT
237 a 315 c 263 g 232 t
ORIGIN
polya_signal
1040..1045
DB
534 TTCACTGAGTTATATGACGACGACTGCTGCTCGACACATGCACTCAAGTCAA 584
QY
1 TTYWNGARTYATGMCARTAYTNGTNTWNSATGCARWSMNCAR 51
RESULT 10
LOCUS HSU48263 1198 bp mRNA PRI 23-AUG-1996
DEFINITION Human pre-pro-orphanin FQ (OFQ) mRNA, complete cds.
ACCESSION U48263
NID 91185009
KEYWORDS
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Carnivora; Homiidae; Homo.
REFERENCE
1 (bases 1 to 1198)
AUTHORS
Nothacker, H.-P., Reinscheld, R.K., Mansour, A., Henningsen, R.A.,
Ardati, A., Monsma, F.J., Jr., Watson, S.J. and Civealli, O.
PRIMARY STRUCTURE AND TISSUE DISTRIBUTION OF THE ORPHANIN FQ
PRECURSOR
PROC. Natl. Acad. Sci. U.S.A. 93 (16), 8677-8682 (1996)
2 (bases 1 to 1198)
Nothacker, H.-P. and Henningsen, R.A.
Direct Submission
Submitted (02-FEB-1996) Hans-Peter Nothacker, PRPN 69/202,
Hoffmann-La Roche AG, Grenzacherstr 124, Basel, 4070, Switzerland
FEATURES
source
Location/Qualifiers
1..1198
/organism="Homo sapiens"
/db_xref="taxon:9606"

```

FEATURES	source	Location/Qualifiers
BASE COUNT	240 a	359 c 341 g 269 t
ORIGIN		
5'UTR		/tissue_type="brain"
gene		1..211
		212..742
		/gene="OFQ"
		212..742
		/gene="OFQ"
		/codon_start=1
		/product="pre-pro-orphanin FQ"
		/db_xref="PID:g1185010"
		/translation="MKVLCDDLILSFSSVSSSCQDCVCRKLRPTLDSFLEWC
		ILDEEKAFPSPLMTPTCTKWARSQWLSPADPHVAALADOPASMOHLKRPVR
		SLDEEVEFPGMEAGEEMQKOLQKFGFTGARSARLANOKRSEFMROYLVS
		MOSSQRRRTIHOGNV"
		212..268
		/gene="OFQ"
		599..649
		/gene="OFQ"
		/product="orphanin FQ"
		743..1198
		/note="29 A nucleotides"
BASE COUNT	244 a	335 c 345 g 273 t 1 others
ORIGIN		
Query Match		92.6%; Score 199; DB 28; Length 1198;
Best Local Similarity		49.0%; Pred. No. 3,656-07;
Matches	25; Conservative	18; Mismatches 8; Indels 0; Gaps 0;
Db	656	TTCACTGAGTTATGAGCATCTGCTGCTCGAGCATGCAGTCACGACG 706
Oy	1	TTWMSGARTTYATGCGNCARTAYTTNGTNTWMSNATGARSWNSNCAR 51
RESULT	11	
LOCUS	AB005251	1209 bp mRNA MAM 09-APR-1998
DEFINITION	Bos taurus mRNA for nociceptin/orphanin FQ precursor, complete cds.	
ACCESSION	AB005251	
NID	g3041768	
KEYWORDS	nociceptin/orphanin FQ precursor.	
SOURCE	Bos taurus	
ORGANISM	Bos taurus	
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Euteria;	
AUTHORS	Artiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.	
TITLE	Okuda-Ashitaka, E., Minami, T., Tachibana, S., Yoshihara, Y.,	
	Nishichii, Y., Kimura, T. and Ito, S.	
	Nocistatin, a peptide that blocks nociceptin action in pain	
	transmission	
JOURNAL	Nature 392 (6673), 286-289 (1998)	
MEDLINE	98180619	
REFERENCE	2 (bases 1 to 1209)	
AUTHORS	Okuda-Ashitaka, E.	
TITLE	Direct Submission	
JOURNAL	Submitted (26-JUN-1997) to the DDBJ/EMBL/GenBank databases. Emko	
	Okuda-Ashitaka, Kansai Medical University, Department of Medical	
	Chemistry, 10-15 Fumizono, Moriyasu, Osaka, JAPAN, Moriyasu,	
	Osaka 570, Japan (E-mail: ashitake@kai.kmu.ac.jp, tel:06-992-1001,	
	Fax:06-992-1781)	
FEATURES		
source		Location/Qualifiers
	1..1209	/organism="Bos taurus"
		/db_xref="taxon:9913"
	210..740	
		/codon_start=1
		/product="nociceptin/orphanin FQ precursor"
		/db_xref="PID:d1026347"
		/db_xref="PID:g3041769"
		/translation="MKILFCDDLILSFSSVSSSCQDCVCRKLRPTLDSFLEWC
		ILDEEKAFPSPLMTPTCTKWARSQWLSPADPHVAALADOPASMOHLKRPVR
		SLDEEVEFPGMEAGEEMQKOLQKFGFTGARSARLANOKRSEFMROYLVS
		MOSSQRRRTIHOGNV"

Query Match 92.6%; Score 199; DB 20; Length 1209;
 Best Local Similarity 49.0%; Pred. No. 3,66e-07;
 Matches 25; Conservative 18; Mismatches 8; Indels 0; Gaps 0;

DB 654 TTCAGTACGTTATGAGCAGTACCGTCCGAGCATGCAAGTCCAGCAG 704
 1 TTYWSNGARTTATGNCARTATYTTNGTNTWSNATGCAWSMNSCAR 51

RESULT 12
 LOCUS MPEPNEX3 1253 bp DNA ROD 10-SEP-1996
 DEFINITION M.musculus gene encoding prepronociceptin, exon 3.
 ACCESSION X97372
 NID 91532038
 KEYWORDS prepronociceptin.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryotes; mitochondria eukaryotes; Metazoa; Chordata;
 Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Mus.

REFERENCE 1 (bases 1 to 1253)
 Mollereau,C., Simons,M.J., Soularue,P., Liners,F., Vassart,G.,
 Meunier,J.C. and Parmentier,M.
 TITLE Structure, tissue distribution, and chromosomal localization of the
 prepronociceptin gene
 Proc. Natl. Acad. Sci. U.S.A. 93 (16), 8666-8670 (1996)
 JOURNAL 96323281
 MEDLINE 2 (bases 1 to 1253)
 AUTHORS Parmentier,M.
 TITLE Direct Submission
 REFERENCE Submitted (19-APR-1996) M. Parmentier, Universite Libre de
 Bruxelles, I R I B H N URB Campus Erasme, 808 Route de Lennik, B-
 1070 Bruxelles, BELGIUM

FEATURES
 source Location/Qualifiers
 1..1253
 /organism="Mus musculus"
 /strain="129/Sv"
 /db_xref="taxon:10090"
 /clone_lib="lambda FIXII (Stratagene #946305)"
 /clone="MG43"
 605..1087
 /number=3
 /label=ex3
 /usedin=X97371:MMPN.CDS
 /usedin=X97371:MMPN.mRNA
 /product="prepronociceptin"
 BASE COUNT 356 a 307 c 323 g 267 t
 ORIGIN

Query Match 92.6%; Score 199; DB 29; Length 1253;
 Best Local Similarity 49.0%; Pred. No. 3,66e-07;
 Matches 25; Conservative 18; Mismatches 8; Indels 0; Gaps 0;

DB 956 TTCAGTACGTTATGAGCAGTACCGTCCGAGCATGCAAGTCCAA 1006
 1 TTYWSNGARTTATGNCARTATYTTNGTNTWSNATGCAWSMNSCAR 51

RESULT 13
 LOCUS MUSNOP 1354 bp mRNA ROD 09-APR-1997
 DEFINITION Mouse mRNA for nociceptin/orphanin FQ, complete cds.
 ACCESSION D82866
 NID 91311472
 KEYWORDS nociceptin/orphanin FQ.
 SOURCE Mus musculus adult brain cDNA to mRNA.
 ORGANISM Mus musculus
 Eukaryotes; mitochondria eukaryotes; Metazoa; Chordata;
 Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae;
 Murinae; Mus.
 REFERENCE 1 (bases 1 to 1354)
 Takeshima,H.
 TITLE Direct Submission

JOURNAL Submitted (25-DEC-1995) to the DBJ/FBI/GenBank databases. Hiroshi
 Takeshima, University of Tokyo, Dept. of Pharmacology, Faculty of
 Medicine, Hongo 7-3-1, Bunkyo-ku, Tokyo 113, Japan
 (E-mail:takeshima@u-tokyo.ac.jp, Tel:03-3812-2111(ex.3422),
 Fax:03-3815-9360)
 REFERENCE 2 (bases 1 to 1354)
 JOURNAL Takeshima,H. (1996)
 AUTHORS Unpublished
 TITLE 3 (sites)
 Houcan,T., Nishi,M., Takeshima,H., Nukada,T. and Sugimoto,T.
 TITLE Structure and regional distribution of nociceptin/orphanin FQ
 precursor
 JOURNAL Biochem. Res. Commun. 219 (3), 714-719 (1996)
 REFERENCE 96216718
 FEATURES
 source Location/Qualifiers
 1..1354
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /dev_stage="adult"
 /tissue_type="brain"
 360..923
 /codon_start=1
 /product="nociceptin/orphanin FQ"
 /db_xref="PID:d1012281"
 /db_xref="PID:g1311473"
 /translation="MKIRFDVLLSLSSVSSCCPDCITCOEKIHPAPDFNLKTC
 SLQVQRAEFGADPEADAEVEYKQLOKRFGEFGARKSARKLANQRF
 SEFRKQYLVLSKQSSQRRRLHONGNV"
 1317..1322
 BASE COUNT 309 a 370 c 386 g 289 t
 ORIGIN

polya.signal
 BASE COUNT 309 a 370 c 386 g 289 t
 ORIGIN

Query Match 92.6%; Score 199; DB 29; Length 1354;
 Best Local Similarity 49.0%; Pred. No. 3,66e-07;
 Matches 25; Conservative 18; Mismatches 8; Indels 0; Gaps 0;

DB 837 TTCAGTACGTTATGAGCAGTACCGTCCGAGCATGCAAGTCAA 887
 1 TTYWSNGARTTATGNCARTATYTTNGTNTWSNATGCAWSMNSCAR 51

RESULT 14
 LOCUS HS1170K4 294095 bp DNA HTG 03-DEC-1998
 DEFINITION Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone
 AL022314
 ACCESSION AL022314
 NID 93947782
 KEYWORDS HTG; HTGS_PHASE1.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
 Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 294095)
 Blakey,S.
 TITLE Direct Submission
 JOURNAL Submitted (03-DEC-1998) Wellcome Trust Genome Campus, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
 On Dec 2, 1998 this sequence version replaced g13927921.
 IMPORTANT: This sequence is unfinished and does not necessarily
 represent the correct sequence. Work on the sequence is in progress
 and the release of this data is based on the understanding that the
 sequence may change as work continues. The sequence may be
 contaminated with foreign sequence from E.coli, Yeast, vector,
 phage etc. Order of segments is not known. 800 's separate
 segments. Unfinished: dj1170K4 Contig_ID: 03636 acc=AL022314
 Length: 20 bp unfinished: dj1170K4 Contig_ID: 02957 acc=AL022314
 Length: 17 bp unfinished: dj1170K4 Contig_ID: 00676 acc=AL022314
 Length: 19 bp unfinished: dj1170K4 Contig_ID: 03380 acc=AL022314
 Length: 19 bp unfinished: dj1170K4 Contig_ID: 04758 acc=AL022314
 Length: 22 bp unfinished: dj1170K4 Contig_ID: 02609 acc=AL022314
 Length: 22 bp unfinished: dj1170K4 Contig_ID: 03488 acc=AL022314


```

Length: 20 bp Unfinished: dj1170K4 Contig_ID: 02907 acc-AL022314
Length: 24 bp Unfinished: dj1170K4 Contig_ID: 01822 acc-AL022314
Length: 19 bp Unfinished: dj1170K4 Contig_ID: 02655 acc-AL022314
Length: 24 bp Unfinished: dj1170K4 Contig_ID: 04353 acc-AL022314
Length: 29 bp Unfinished: dj1170K4 Contig_ID: 03931 acc-AL022314
Length: 20 bp Unfinished: dj1170K4 Contig_ID: 01036 acc-AL022314
Length: 21 bp Unfinished: dj1170K4 Contig_ID: 00860 acc-AL022314
Length: 28 bp Unfinished: dj1170K4 Contig_ID: 02949 acc-AL022314
Length: 20 bp Unfinished: dj1170K4 Contig_ID: 00341 acc-AL022314
Length: 42 bp Unfinished: dj1170K4 Contig_ID: 03201 acc-AL022314
Length: 33 bp Unfinished: dj1170K4 Contig_ID: 04365 acc-AL022314
Length: 17 bp Unfinished: dj1170K4 Contig_ID: 05023 acc-AL022314
Length: 3767 bp Unfinished: dj1170K4 Contig_ID: 01947
acc-AL022314 Length: 18 bp Unfinished: dj1170K4 Contig_ID: 01270
acc-AL022314 Length: 19 bp Unfinished: dj1170K4 Contig_ID: 01583
acc-AL022314 Length: 18 bp Unfinished: dj1170K4 Contig_ID: 03812
acc-AL022314 Length: 29 bp Unfinished: dj1170K4 Contig_ID: 04647
acc-AL022314 Length: 26 bp Unfinished: dj1170K4 Contig_ID: 02527
acc-AL022314 Length: 21 bp Unfinished: dj1170K4 Contig_ID: 03783
acc-AL022314 Length: 25 bp Unfinished: dj1170K4 Contig_ID: 01804
acc-AL022314 Length: 20 bp Unfinished: dj1170K4 Contig_ID: 04857
acc-AL022314 Length: 33 bp Unfinished: dj1170K4 Contig_ID: 03559
acc-AL022314 Length: 18 bp Unfinished: dj1170K4 Contig_ID: 00485
acc-AL022314 Length: 19 bp Unfinished: dj1170K4 Contig_ID: 02367
acc-AL022314 Length: 25 bp Unfinished: dj1170K4 Contig_ID: 02270
acc-AL022314 Length: 23 bp Unfinished: dj1170K4 Contig_ID: 00018

```

*** WARNING: Phase 1 High Throughput Genome Sequence ***

* This sequence data is unfinished. When sequencing is complete,
* the sequence data presented in this record will be replaced.
* by a single finished sequence with the same accession number.

FEATURES
source
1. 294095
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/clone="1170K4"

BASE COUNT 64545 a 71590 c 69905 g 62435 t 25620 others

ORIGIN

Query Match 57.7% Score 124 DB 18 Length 294095;

Best Local Similarity 37.5% Pred. No. 6.77e+00;
Matches 18; Conservative 16; Mismatches 14; Indels 0; Gaps 0;

Db 272427 CTCAGGCTCATTAACGCAATACCTGCGCTGATGATGACAGATGACA 272474

Oy 3 YWSNCAATTATGNCNCARATAYTNGTNTNMSNATGCAFMNSMCA 50

RESULT 15
LOCUS AC002541 166983 bp DNA PRI 12-SEP-1997
DEFINITION Human BAC clone RG043K06 from 7q21-q22, complete sequence.
AC002541
ACCESSION 92393735
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 166983)
Minx,P., Tin,A., Sutterer,C and Hawkins,M.
The sequence of H. sapiens BAC clone RG043K06
Unpublished (1997)
2 (bases 1 to 166983)
Waterston,R.
Direct Submission
Submitted (12-SEP-1997) Department of Genetics, Washington
University, 444 Forest Park Avenue, St. Louis, Missouri 63108, USA
SUBMITTED BY:
Genome Sequencing Center
Department of Genetics

Washington University
St. Louis MO 63108, USA
http://genome.wustl.edu/gsc
mailto:sapiens@wustl.wustl.edu

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see
http://www.nhgri.nih.gov/DIR/CTB/CHR7 or send
mailto:egreen@nhgri.nih.gov

SOURCE INFORMATION:
This clone is from the first release of the human BAC library. The library contains cloned DNA from a human male fibroblast cell line 978SK. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 88:8794-7 (1992); Kim et al., Genomics 34:213-8 (1996). The clone is available from Research Genetics, Inc. (http://www.resgen.com).
VECTOR: pBelobAC11
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:
The actual start of this clone is at base position 1 of RG043K06; actual end is at 166983 of RG043K06. The orientation of this clone is unknown.

FEATURES
This clone contains STS SWSS1326 (NTD:940334).

source
Location/Qualifiers
1. 166983
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/clone="RG043K06"
/map="7q21-q22"

repeat_region 650..703
/rpt_family="L1"
repeat_region 1026..1751
/rpt_family="L1"
repeat_region 2172..5929
/rpt_family="L1"
complement(4862..5266)
/rpt_family="L1"
repeat_region 6072..6164
/rpt_family="L1"
repeat_region 6304..6584
/rpt_family="L1"
repeat_region 7600..7637
/rpt_family="L1"
repeat_region 11880..11901
/rpt_family="L1"
repeat_region 11978..12442
/rpt_family="L1"
repeat_region 12523..12812
/rpt_family="ALU"
complement(20178..20466)
/rpt_family="ALU"
repeat_region 23366..24826


```

repeat_region /rpt_family="L1" 26530. .26570
repeat_region /rpt_family="L1" complement(30991. .31369)
repeat_region /rpt_family="L1" 30991. .31392
repeat_region /rpt_family="L1" 34123. .34162
repeat_region /rpt_family="L1" complement(34281. .34405)
repeat_region /rpt_family="L1" complement(35591. .35645)
repeat_region /rpt_family="L1" 38226. .38311
repeat_region /rpt_family="L1" 39072. .39461
repeat_region /rpt_family="L1" complement(40536. .41041)
repeat_region /rpt_family="L1" 43083. .43590
repeat_region /rpt_family="L1" 44156. .44186
repeat_region /rpt_family="L1" complement(44592. .44618)
repeat_region /rpt_family="L1" 45084. .45112
repeat_region /rpt_family="L1" 47602. .47635
repeat_region /rpt_family="L1" 49901. .50192
repeat_region /rpt_family="ALU" 50767. .51049
repeat_region /rpt_family="ALU" complement(53778. .53816)
repeat_region /rpt_family="L1" 55314. .55367
repeat_region /rpt_family="L1" 55997. .57239
repeat_region /rpt_family="L1" complement(57235. .57272)
repeat_region /rpt_family="L1" 57271. .57305
repeat_region /rpt_family="L1" 57853. .58150
repeat_region /rpt_family="ALU" 58899. .58930
repeat_region /rpt_family="L1" complement(58936. .59015)
repeat_region /rpt_family="L1" complement(59214. .59271)
repeat_region /rpt_family="L1" 60343. .60364
repeat_region /rpt_family="L1" complement(61043. .61247)
repeat_region /rpt_family="L1" 61043. .61248
repeat_region /rpt_family="L1" 64833. .64870
repeat_region /rpt_family="L1" 66745. .66929
repeat_region /rpt_family="L1" 68562. .68595
repeat_region /rpt_family="L1" complement(68602. .68886)
repeat_region /rpt_family="ALU" 69060. .69201
repeat_region /rpt_family="L1" complement(69202. .69416)
repeat_region /rpt_family="ALU" 69418. .70100
repeat_region /rpt_family="L1" complement(69534. .69940)
repeat_region /rpt_family="L1"

```

```

repeat_region 70118. .70278
repeat_region /rpt_family="L1" complement(70239. .70749)
repeat_region /rpt_family="L1" 70320. .71433
repeat_region /rpt_family="L1" complement(71540. .71685)
repeat_region /rpt_family="MER" complement(71880. .72079)
repeat_region /rpt_family="MER" complement(72426. .72561)
repeat_region /rpt_family="MER" 72472. .72930
repeat_region /rpt_family="MER" 73073. .73174
repeat_region /rpt_family="MER" 73179. .73468
repeat_region /rpt_family="ALU" 73488. .73561
repeat_region /rpt_family="MER" 73562. .73605
repeat_region /rpt_family="L1" 73606. .73766
repeat_region /rpt_family="MER" 74103. .74358
repeat_region /rpt_family="L1" 74492. .75185
repeat_region /rpt_family="L1" complement(77916. .78147)
repeat_region /rpt_family="ALU" 78284. .78313
repeat_region /rpt_family="L1" complement(79753. .79791)

```

*** Note: remainder of annotations omitted.

```

Query Match 55.3%; Score 119; DB 27; Length 166983;
Best Local Similarity 36.7%; Pred. No. 1.83e+01;
Matches 18; Conservative 16; Mismatches 15; Indels 0; Gaps 0;
DB 79936 GGCATCATGACCTTTTAAGCCCAATACTTAAGCATAACTGGAAN 79984
CP 49 GNSMNSWITGCATNSWMAACNARATAYTGCKCATRAATGNSWRA 1

```

Search completed: Tue Apr 27 10:23:42 1999
Job time : 146 secs.

 WIRENET (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPerch_tpn n.a. - n.a. Smith-Waterman search, using a protein query
 which has been backtranslated into n.a. using IUPAC symbols

Run on: Tue Apr 27 10:26:44 1999; MasPar time 25.91 Seconds

Tabular output not generated. 267.841 Million cell updates/sec

Title: >US-09-011-797-3
 Description: (1-17) from US09011797.pep
 Perfect Score: 215
 N.A. Sequence: 1 TTYWSNGARTTYATGNGNCA.....TMSNATGCAKMSNSNCAR 51
 Comp: AARMSNCTYAARTACKNCGT.....AMWSNACTGYMSNGNGTY

Scoring table: TABLE bkxtranslated2
 Gap 40

Nmatch STD : Dbase 0; Query 0

Searched: 188442 seqs, 68026449 bases x 2

Post-processing: Minimum Match 08
 Listing first 45 summaries

Database: n-geneseq32
 1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
 8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
 14:part14 15:part15 16:part16 17:part17 18:part18
 19:part19 20:part20 21:part21 22:part22 23:part23
 24:part24 25:part25 26:part26 27:part27 28:part28
 29:part29 30:part30 31:part31 32:part32 33:part33
 34:part34 35:part35 36:part36 37:part37 38:part38
 39:part39 40:part40

Statistics: Mean 37.150; Variance 152.162; scale 0.244

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	199	92.6	932	34	T79888	DNA encoding human op	3.11e-07
2	108	50.2	1151	4	Q26021	PSP6-SFV4 RNA transcr	1.18e+01
3	106	49.3	1873	13	Q80220	Human NDF-beta1a c1on	1.67e+01
4	105	48.8	988	31	T43486	ARM gene exons 53-54.	1.98e+01
5	104	48.4	271	2	M60950	Derivative of plasmid	2.34e+01
6	103	47.9	3016	19	T17115	Rhodopsin gene.	2.78e+01
7	102	47.4	2504	10	Q90443	Murine oviduct specif	3.29e+01
8	99	46.0	405	10	Q44373	Sequence of VHH DNA s	5.45e+01
9	98	45.6	315	8	O60950	Human brain Expressed	6.44e+01
10	98	45.6	1692	1	O03702	Gene encoding plant e	6.44e+01
11	98	45.6	2301	22	T32324	Soluble starch syntha	6.44e+01
12	98	45.6	3314	10	Q58896	Humanised anti-CEA sF	6.44e+01

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
13	98	45.6	3540	2	O11854	Glutamate receptor 6.	6.44e+01
14	98	45.6	4127	31	T86846	Potato soluble starch	6.44e+01
15	97	45.1	423	33	T87071	Type II topoisomerase	7.60e+01
16	97	45.1	540	11	O62072	L. heterohlochi JCM11	7.60e+01
17	97	45.1	540	15	O88205	Lactobacillus sp. 165	7.60e+01
18	97	45.1	1133	36	T74986	Human H1A DR alpha te	7.60e+01
19	97	45.1	1972	8	Q49050	Fungus-responsive pot	7.60e+01
20	97	45.1	5737	17	T12235	pp60Pik cDNA.	7.60e+01
21	96	44.7	2735	2	O11072	B.steiothermophilus	8.97e+01
22	95	44.2	356	6	O39131	Mouse MK coding seque	1.06e+02
23	95	44.2	440	36	T88799	Nuclear steroid hormo	1.06e+02
24	94	43.7	1365	1	O04715	Human arginase cDNA c	1.24e+02
25	94	43.7	1407	15	O92081	Human pexoxisome proI	1.24e+02
26	94	43.7	3614	35	T91883	Murine SIM (single-mi	1.24e+02
27	94	43.7	3981	1	N70695	Macrophage-colony sti	1.24e+02
28	94	43.7	3981	1	N91734	Sequence of a macroph	1.24e+02
29	94	43.7	4021	2	O10343	Monocyte macrophage c	1.24e+02
30	94	43.7	7454	20	T29660	Nisin nisAaRCP gene	1.46e+02
31	93	43.3	274	34	T61624	Streptococcus thermop	1.46e+02
32	93	43.3	403	8	O60838	Human brain Expressed	1.46e+02
33	93	43.3	413	40	V16480	DNA encoding a Bac111	1.46e+02
34	93	43.3	452	20	T22420	Human gene signature	1.46e+02
35	93	43.3	652	5	O34564	Shrgp.	1.46e+02
36	93	43.3	1313	27	T48101	Immunogenic type F bo	1.46e+02
37	93	43.3	1449	32	T72162	N-acetylglucosamine 1	1.46e+02
38	93	43.3	2799	29	T42854	Phospholipase D gene	1.46e+02
39	93	43.3	2799	16	O86785	DNA encoding Phosphol	1.46e+02
40	93	43.3	3075	37	V06252	Murine purmycin-sens	1.46e+02
41	93	43.3	3695	6	O32844	Encodes E. coli heat	1.46e+02
42	93	43.3	3762	6	O38738	Encodes E. coli heat	1.46e+02
43	93	43.3	5579	23	T30056	S. pneumoniae detecti	1.46e+02
44	93	43.3	6953	19	T17116	Rhodopsin gene.	1.46e+02
45	92	42.8	8585	2	N60404	Sequence encoding hum	1.72e+02

ALIGNMENTS

RESULT 1
 ID T79888 standard; DNA; 932 BP.
 AC T79888;
 DT 08-DEC-1997 (first entry)
 DE DNA encoding human opiod receptor-like 1 receptor ligand.
 KW Opiod receptor-like 1; ORL1; ligand; pronociceptive; stress;
 KW hyperalgesia; locomotor activity disease; anxiety; neuroendocrine;
 KW memory; attention; sensory perception; learning; homeostasis;
 OS Synthesis.
 FH Key Location/Qualifiers
 FT cds 1..507
 FT /tag= a
 FT /note= "no start codon"
 PN WO9707208-A1.
 PD 27-FEB-1997.
 PF 14-AUG-1996; BE0087.
 PR 15-AUG-1995; US-002368.
 PA (ULBR) UNIV LIBRE BRUXELLES.
 PI Meunier J, Molleau C, Parmentier M, Vassart G;
 DR WPI: 97-165292/15.
 DR P-PSDB: W25162, W25163, W25164.
 PT Novel ligand for the opiod receptor-like receptor, nociceptin -
 PT has pronociceptive properties, useful for treating or preventing
 PT diseases related to e.g. stress, hyperalgesia, locomotor activity,
 PT etc.
 PS Claim 1: Page -? 48pp; English.
 CC T79888 encodes a ligand to human opiod receptor-like 1 (ORL1) receptor
 CC designated nociceptin, which resembles the endorphin dynorphin A.
 CC Inhibitors of nociceptin or its DNA, e.g. antisense sequences and
 CC antibodies can be used as new types of drugs in the control of various
 CC behaviours or functions. The inhibitors can be used to prevent or treat
 CC a disease related to hyperalgesia, neuroendocrine secretion, stress,
 CC locomotor activity, anxiety, instinctive behaviour, learning and memory,
 CC homeostasis, hyperalgesia, hypalgesia and/or sensory perception.
 SQ Sequence 932 BP; 224 A; 273 C; 238 G; 197 T;

KW	eye; rod; retina; diagnostic; prenatal diagnosis; photoreceptor; ds
OS	Homo sapiens.
FH	key location/Qualifiers

DT	10-JAN-1996
DE	Murine ovid
KW	Murine ovid

KM MOGP: ds.
OS Mus musculus.
FH Key Location/Qualifiers
1 3..2160
FT signal_peptide 3..56 /note= "START codon absent"
FT mat_peptide 57..2156 /tag= b
FT /tag= c
PN J07107979-A.
PD 25-APR-1995.
PF 15-AUG-1994; 214227.
PR 19-AUG-1993; JP-227881.
PA (KINO-) KINOSRI PEPTIDE KENKYUSHO KK.
DR WPI: 95-190179/25.
DR P-PSDB: R90443.
PT New DNA encoding an oviduct-specific glycoprotein - useful for
PS recombinant protein production in high quantities.
CC Claim 1; Pages 11-14; 22pp; Japanese.
CC Q90443 encodes R73992 murine oviduct specific glycoprotein (MOGP).
CC The cDNA can be used for the commercial recombinant prodn. of
CC MOGP in high quantities.
SQ Sequence 2504 BP; 678 A; 616 C; 538 G; 672 T;
Query Match 47.4%; Score 102; DB 15; Length 2504;
Best Local Similarity 38.5%; Pred. No. 3.29e+01;
Matches 15; Conservative 12; Mismatches 12; Indels 0; Gaps 0;
Db 1198 atgcctcagcagcctcctcgaatacagcctcagc 1236
Oy 13 ATGMCARTATYATGNTWMSNATGCARFWSMNCAR 51
RESULT 8
AC Q44373 standard; DNA; 405 BP.
ID Q44373:
DT 22-AUG-1994 (first entry)
DE Sequence of VHH DNA sequence DR07006 of camel heavy chain
DE Immunoglobulin (Ig).
KW Immunoglobulin; heavy chain; Camelid.
OS Camelus dromedarius.
PN MO9404678-A.
PD 03-MAR-1994.
PF 18-AUG-1993; E02214.
PR 21-AUG-1992; EP-402326.
PR 21-MAY-1993; EP-401310.
PA (CAST/) CASTERMAN C.
PA (HAME/) HAMERS R.
PI Casterman C. Hamers R.
DR WPI: 94-083195/10.
PT Immunoglobulins devoid of light chains - also processes for their
PS preparation, and protein and nucleotide sequence encoding them.
CC Claim 11; Figure 7; 87pp; English.
CC A novel immunoglobulin (Ig) is claimed which comprises two heavy (H)
CC polypeptide chains sufficient for the formation of a complete
CC antigen binding site or several such chains. The Ig is devoid of
CC light (L) polypeptide chains. The Ig may be obtd. from prokaryotic
CC cells, esp. E. coli, by: cloning a DNA or cDNA sequence coding for
CC the VH domain of an Ig devoid of L chains obtainable from e.g.
CC lymphocytes of Camelids; recovering the cloned fragment after
CC amplification using a 5' primer contg. an xho site and a 3' primer
CC contg. the Spe site having the sequence in Q44383; cloning the
CC recovered fragment is a vector; transforming host cells; and
CC recovering the expression product of the VHH coding sequence.
CC A nucleotide sequence encoding the novel Ig, selected from
CC Q44366-43382, is claimed.
SQ Sequence 405 BP; 92 A; 99 C; 114 G; 100 T;
Query Match 46.0%; Score 99; DB 10; Length 405;
Best Local Similarity 33.3%; Pred. No. 5.45e+01;
Matches 17; Conservative 16; Mismatches 18; Indels 0; Gaps 0;

Db 198 ctcccaagacagcagcgttgagacagatgctcgtcaatgaacacccgaa 248
Oy 51 YTGSMNSWYTGCAITNSWNRNACNRRTATYGNCGCAITRAITNSWRA 1
RESULT 9
AC Q60950 standard; DNA; 315 BP.
ID Q60950:
DT 16-MAR-1994 (first entry)
DE Human brain Expressed Sequence Tag EST02695.
KW Gene transcription product; genetic markers; tagging; in vivo;
OS Homo sapiens.
PN MO9316178-A.
PD 19-AUG-1993.
PF 12-FEB-1993; U01294.
PR 12-FEB-1992; US-837195.
PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
PI Adams MD, Moreno RF, Venter CJ.
DR WPI: 93-272882/34.
PT Enriched oligonucleotides and corresp. sequences - used as
PT markers for human genes transcribed in-vivo, facilitate tagging
PS Example 4; Page 416; 500pp; English.
CC The Expressed Sequence Tag was isolated from a human brain cDNA
CC library as part of a large set of ESTs which can be used as markers
CC for human genes transcribed in vivo. They can be used to facilitate
CC tagging of most human genes, for mapping locations of expressed genes
CC on chromosomes, for individual or forensic identification, for mapping
CC locations of disease-associated genes, for identification of tissue
CC type, and for prep. of antisense sequences, probes and constructs.
CC EST02695 has a "poor" coding probability as evaluated using the
CC coding-region prediction program CRM. See also Q59041-Q61440.
SQ Sequence 315 BP; 91 A; 43 C; 56 G; 124 T;
Query Match 45.6%; Score 98; DB 8; Length 315;
Best Local Similarity 38.6%; Pred. No. 6.44e+01;
Matches 17; Conservative 13; Mismatches 14; Indels 0; Gaps 0;
Db 119 ttgatagaattgttgatgcgattgtgctcttgattcagtc 162
Oy 1 TTWMSNCARTATYATGMCARTATYATGNTWMSNATGCARFWS 44
RESULT 10
AC Q03702 standard; DNA; 1692 BP.
ID Q03702:
DT 06-AUG-1990 (first entry)
DE Gene encoding plant elongation factor.
KW Plant elongation factor-1 alpha; ds.
OS Lycopersicon esculentum.
FH Key Location/Qualifiers
FT cds 61..2004 /tag= a
FT MO9002172-A.
PN 8-MAR-1990.
PD 17-AUG-1989; 003536.
PF 18-AUG-1988; US-234187.
PR 16-AUG-1988; US-335133.
PR 07-MAR-1989; US-335133.
PA (CALC-) Calgene Inc.
PI Shewmaker CK, Hiatt WR, Pokalsky AR;
DR WPI: 90-099398/13.
DR P-PSDB: R05565.
PT Plant elongation factor -
PT has plant promoter allowing elevated expression in rapidly
PT dividing cells.
PS Claim 18; Page 58; 80pp; English.
CC Transcription regulatory regions provide a means of expression in rapidly
CC dividing tissue with lower levels of expression in mature cells. The
CC gene may thus be used to regulate expression in transgenic plants.
CC Gene also provides for identification of other similar conserved
CC sequences associated with plant elongation factor.
SQ Sequence 1692 BP; 399 A; 351 C; 430 G; 512 T;


```

Query Match          45.6% Score 98; DB 1; Length 1692;
Best Local Similarity 40.6%; Pred.No. 6,44e+01;
Matches 13; Conservative 11; Mismatches 8; Indels 0; Gaps 0;

Db    1540 ctgcacagatcatcttcacagaactggttact 1571
      :||:::|||||:::|:||::|||:|
Cp    51 ytgmswmwtygcattinswmnarnactrratt 20

RESULT 11
ID T33324 standard; cDNA; 2301 BP.
AC T33324;
DT 30-OCT-1996 (first entry)
DE Soluble starch synthase (isoform A) coding sequence.
KW Soluble starch synthase; transformation; transgenic plant; starch;
   amylose; amylopectin; ss.
OS Solanum tuberosum.
FT Key Location/Qualifiers
FT cds 1..2036
FT     /*tag= a
FT     /product= Soluble starch synthase.
PN DE4441408-A1.
PF 15-MAY-1996.
PR 10-NOV-1994; 441408.
PA (GENB-) INST GENBIOLGISCHE FORSCHUNG.
PI Abel GJ, Kossmann J, Springer F;
DR WPI: 96-240218/25.
PT P-PSDB: R895359.
PT DNA encoding soluble starch synthase of potato - used to produce
PT transgenic plants with increased prodn. of starch or able to produce
PT modified starches
PS Claim 1; Page 16-19; 32pp; German.
CC DNA sequences from the potato encoding soluble starch synthase
CC having the 677 C-terminal amino acid sequence encoded by this
CC sequence or the 459 C-terminal amino acid sequence given in R89540
CC can be used to identify and isolate homologous sequences encoding
CC soluble starch synthase and enzymes with similar activities from
CC plants or other organisms; to transform prokaryotic or eukaryotic
CC cells; to produce transgenic plants which synthesise starch of
CC altered structure or in increased yield.
SV Sequence 2301 BP; 666 A; 435 C; 522 G; 678 T;

Query Match          45.6% Score 98; DB 22; Length 2301;
Best Local Similarity 37.8%; Pred.No. 6,44e+01;
Matches 17; Conservative 13; Mismatches 15; Indels 0; Gaps 0;

Db    1775 gaccttatgatcgtattgatgttcgaccatgacaagaagagag 1819
      |||||:::|:||::|||::|:||::|||:::|:|
Oy    7 GARTTYATGMGNCAATAYTNGTNTYTNMNSNATGCARMSNCAR 51

RESULT 12
ID O58896 standard; DNA; 3314 BP.
AC O58896;
DT 26-OCT-1994 (first entry)
DE Humanised anti-CEA sFv fragment-human beta-glucuronidase fusion gene.
KW Carcinoembryonic antigen; single chain variable region; sFv fragment;
   fusion gene; cancer treatment; targeted drug delivery; tumour; ss.
OS Homo sapiens.
FT Key Location/Qualifiers
FT exon 1..189
FT     /*tag= a
FT     /codon_start= 145
FT     /note= "encodes amino acids -19 to -5"
FT     272..1003
FT     /*tag= b
FT     /number= 2
FT     /note= "encodes amino acids -4 to 240"
FT     1069..1119
FT     /*tag= c
FT     /number= 3

```

[illegible]


```

RESULT 14
ID T68646 standard; CDNA: 4127 BP.
AC T68646;
DT 16-AUG-1997 (first entry)
DE Potato soluble starch synthase cDNA.
KW Starch synthase; transgenic plant; potato; rice; Oryza sativa;
KW tomato; Lycopersicon esculentum; wheat; Triticum aestivum; cassava;
KW Manihot esculenta; sweet potato; Ipomoea batatas; barley;
KW Hordeum vulgare; oat; Avena; maize; Zea mays; ds.
OS Solanum tuberosum cv. Desiree.
FH Key Location/Qualifiers
FT cds 143..3835
FT signal_peptide 143..322
FT mat_peptide 323..3832
FT /tag= b
FT /tag= c
PN EP-779363-A2.
PD 18-JUN-1997.
PF 11-DEC-1996; 309004.
PR 12-DEC-1995; GB-025353.
PA (NAT) NAT STARCH & CHEM INVESTMENT HOLDING COR.
PI Edwards EA, Marshall J, Martin CR, Smith AM,
DR MPI; 97-312737/29.
DR P-PSDB; W17765.
PT Soluble starch synthase - used to produce altered starch from
PT commercially important plants, e.g. potato, rice, wheat, and maize
PS Claim 16; Page 18-24; 39pp; English.
CC A composite cDNA sequence (T68646) codes for a 100-140 kDa potato
CC soluble starch synthase (W17765). The sequence was deduced from 2
CC overlapping cDNA clones obtd. from a tuber cDNA library by
CC immunoscreening with antiserum raised against purified starch
CC synthase. Starch synthase nucleic acids can be linked to a plant
CC promoter, in sense or antisense direction, and used to alter starch
CC characteristics in transgenic plants, esp. potato, tomato, rice,
CC wheat, pearl cassava, sweet potato, barley, oat or maize. A novel
CC altered starch, extracted from such transgenic plants, has a
CC viscosity onset temperature that is reduced by at least 5 degC
CC compared to starch from non-transformed plants.
SQ Sequence 4127 BP; 1282 A; 709 C; 977 G; 1159 T;

Query Match 45.6%; Score 98; DB 31; Length 4127;
Best Local Similarity 37.8%; Pred. No. 6.44e+01;
Matches 17; Conservative 13; Mismatches 15; Indels 0; Gaps 0;

Db 3576 gacttatgatacgtattgattgacatgacaaagagagag 3620
OY 7 GARTTYATGMCNARTAYTNGTNTWMSNATGCARWSNMCAR 51

RESULT 15
ID T80701 standard; DNA: 423 BP.
AC T80701;
DT 11-NOV-1997 (first entry)
DE Type II topoisomerase database reference sequence SEQ ID NO:120.
KW Detection; identification; TopoII; contamination; food; human;
KW animal; plant; soil; water; PCR; polymerase chain reaction; ds.
OS Borrelia garinii.
PN US5645994-A.
PD 08-JUL-1997.
PF 05-JUL-1990; 548138.
PR 06-JUN-1995; US-470179.
PR 05-JUL-1990; US-548138.
PR 13-AUG-1993; US-106482.
PA (UTAH) UNIV UTAH RES FOUND.
PI Huang WM.
DR MPI; 97-362925/33.
PT Detection and identification of organisms - using DNA primers to
PT amplify signature segment of organism's type II topoisomerase
PS Claim 6; Column 121-124; 114pp; English.
CC A method has been produced for selectively amplifying DNA segments of

```

```

CC one or more species of organisms in a sample. The method involves: (a)
CC providing a database containing reference sequences, comprising a
CC subunit sequence of a signature region of a macromolecule selected from
CC a type II topoisomerase (TopoII), or a homologue, where each reference
CC sequence is specific to a different species of a chosen group, and the
CC macromolecule comprises 1st and 2nd conserved regions adjacently
CC flanking the signature region; and (b) making an extract of DNA
CC molecules, and selectively amplifying DNA segments of the signature
CC region using a universal primer composition, comprising a primer
CC constructed to bind a DNA encoding the macromolecule, to produce
CC amplified DNA segments. The present sequence represents a DNA fragment
CC containing a signature segment which is used in the database as a
CC reference sequence for Borrelia garinii. The method can be used to
CC identify all of the different organisms present in a single
CC sample without using multiple probes. It can accurately distinguish
CC among similar and related species, and can be used with, e.g.
CC contaminated food products, tissue or fluid samples from diseased
CC humans, animal or plants, soil samples and water samples from any
CC source.
SQ Sequence 423 BP; 125 A; 64 C; 94 G; 140 T;

Query Match 45.18; Score 97; DB 33; Length 423;
Best Local Similarity 41.7%; Pred. No. 7.60e+01;
Matches 15; Conservative 11; Mismatches 10; Indels 0; Gaps 0;

Db 16 gtccatagagagaatacttattctctatgataag 51
OY 9 RTTYATGMCNARTAYTNGTNTWMSNATGCARWS 44

Search completed: Tue Apr 27 10:27:17 1999
Job time : 33 secs.

```


Contact: Marra M/Mouse EST Project

Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LINT ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI: 203590

Seq primer: mob.REGA+ET
High quality sequence stop: 382
Location/Qualifiers
1..388

/organism="Mus musculus"
 /note="Vector: pT7TD (Pharmacia) with a modified
 polylinker: Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
 was primed with a Not I - oligo(dT) primer [5',
 TGTTACCAATCTAGAGGAGGAGGCGCCATTTTCTTTTCTTTT
 3'],
 double-stranded cDNA was size selected, ligated to Eco RI
 adapters (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of a modified pT73 vector
 (Pharmacia). Library went through one round of
 normalization to a Cot = 5. Library constructed by Bento
 Soares and M. Fatima Bonaldo. RNA was kindly provided by
 Dr. Minoru Ko (Wayne State University)."
 /db="GenBank" /accession="U00000" /version="1.0"

BASE COUNT	100 a	98 c	96 g	94 t
ORIGIN				

52.68; Score 113; DB 10; Length 388;

```
18; Conservative 12; Mismatches 12; Indels 0; Gaps 0;
```

Db 6 TGGAAGGATGATGCATCTGTGTACACAGAAATTGTACATGTAAC 47
|| : :: ||| : : || : || : ||| : ||
Cp 50 TCGSNWSWTGCATSNWNRNRCNARRATYTGCKCATTAAY 9

RESULT	5				
LOCUS	AA796248	421 bp	mRNA	EST	09-FEB-1998
DEFINITION	v998c09.r1 Barestead mouse myotubes MPLRB5 Mus musculus cDNA clone				
Accession	1154320.5	mRNA sequence.			

NID	92859203
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus

Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria
Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE	AUTHORS
1 (bases 1 to 421)	
	Marta, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.

TITLE	COMMENT
The WashU-HHMI Mouse EST Project	
Unpublished (1996)	

Contact: Maria M/Mouse EST Project
WASHU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNNI; con

Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 419.

FEATURES
SOURCE

```

/organism="Mus musculus"
/strain="C3H"
/note="Vector: pT73D-Pac (Phar.McA.2: NotI; 1st strand cDNA
polylinker; Site.1: EcoRI; Site.2: NotI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTGACGACATCTGAAGGGAGGAGCGCCCTCTTTTCTTTTCTTTTCTTTT
3'] ; double-stranded cDNA was ligated to Eco RI adaptors
[AAATCGACATCTG], digested with Not I and cloned into the
Not I and Eco RI sites of the modified pT73 vector.
The Not I and Eco RI sites of the modified pT73 vector.
Library constructed by Bob Bartstead. The C612 cell line
clonable from ATCC, catalog # CRL-1172) differentiates
rapidly, forming contractile myotubes and producing
characteristic muscle proteins."
/db_xref="taxon:10090"

```

BASE COUNT	103 a	108 c	106 g	104 t
ORIGIN				

Query Match	Score	DB	Length
52.68;	113;	12;	421

Best Local Similarity 42.96; Pred.NO. 3.28E-03;
Matches 18; Conservative 12; Mismatches 12; Indels 0; Gaps 0.

Db 52 TGAAGGATAGCATCTCTGTGTACAGAATTGTCACATGAC 93

50 TGN^{SWNS}WYTCATNSW^{NAR}NACNAR^{TAY}TGNC^{KCA}TRAY 5

LOCUS	6	RESULT
DEFINITION	AA796116 470 bp mRNA EST 09-FEB-1998	
DEFINITION	v066e06.1 Soares mouse mammary gland NBWMC Mus musculus cDNA clone	
DEFINITION	1054882 5', mRNA sequence.	
DEFINITION	AA706116	

NID	g2859071
KEYWORDS	EST.
SOURCE	house mouse

ORGANISM
Mus musculus
Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria

REFERENCE
AUTHORS
1 (bases 1 to 470)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.

Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

TITLE The WashU-HHMI Mouse Est Project
JOURNAL Unpublished (1996)
COMMENT

Contact: Marra M/Mouse EST Project
 Mashu-HHM Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63105
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LNL ; contact
 IMAGE Consortium (lnl@image.llnl.gov) for further info
 MGI:566458
 Seq primer: -28m13 rev2 ET from Amer sham.
 Location/Qualifiers

```

source
1..470
/organism="Mus musculus"
/strain="C57Bl/6J"
/note="Organ: mammary gland; Vector: pT73D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I

```


Query Match	52.6%	Score 113	DB 12	Length 470
Best Local Similarity	42.9%	Pred. No.	3.28e-03	
Matches	18	Conservative	12	Mismatches 12; Indels 0; Gaps 0;
Db	87	TGGAGGATGCACTCCTCGTGTACAGAAATTTGTCACTGAAC	128	
	. : : :	.. : : : : : : : : : : :		
Cp	50	TGNSMNSWITGCTCATNSMNAANNCNARFATYTCNCCATPAAAY	9	

RESULT	7	AA269345	473 bp	MRNA	EST	26-MAR-1997
LOCUS		vb09g05.r1	Soares mouse	NML	Mus musculus	cDNA clone.748472.5', MRNA
DEFINITION		sequence.				
ACCESSION		AA269345				
NID		91907624				
KEYWORDS		EST.				
SOURCE		house mouse.				
ORGANISM		Mus musculus				
		Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE		1 (bases 1 to 473)				
AUTHORS		Merra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,U., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.				
TITLE		The WashU-HHMI Mouse EST Project				
JOURNAL		Unpublished (1996)				
COMMENT						

TITLE	JOURNAL	COMMENT
The WashU-HHMI Mouse EST Project	Unpublished (1996)	Contact: Marra M/Mouse EST Project WashU-HHMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LINT; contact the IMAGE Consortium (info@image.lnl.gov) for further information. MGI:457456 Seq primer: -28m13 rev2 ET from Amersham High quality sequence stop: 466. Location/Qualifiers 1..473

```

/organism="Mus musculus"
/notes="Vector: pT7SD-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTGCAATCTGAAGGCGGCGCGGATTTTTTTTTTTTTTTT 3'];
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
constructed and normalized by Bento Soares and M.Fátima
Donaldso."

```

RESULT	8								
LOCUS	AA795004	488 bp	mRNA	EST	09-FEB-1998				
DEFINITION	vso8c07.t1 Barstead mouse irradiated colon NFE1B7 Mus musculus cdNNA								
ACCESSION	AA795004								
NID	62857955								
KEYWORDS	EST.								
SOURCE	house mouse.								
ORGANISM	Mus musculus								

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (pages 1 to 488)	Marr, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Knaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Wreising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.	The WASHU-HHMI Mouse EST Project	Unpublished (1996)	

FEATURES
 source
 Contact: Marra M/Mouse EST Project
 WashU-HHMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel.: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.edu
 This clone is available royalty-free through INLNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 M01:418884
 Putative full length read
 vector to vector length is 564
 Seq primer: -28ml3 rev2 ET from Amersham
 High quality sequence stop: 463.
 Location/Qualifiers
 1..488

```

BASE COUNT      125 a      121 c      120 g      122 t
ORIGIN
/clone="p17730-Pac (Pharmacia) with a modified
polylinker; Site_1: EcoRI; Site_2: NotI; Tissue obtained
from 8 week old mouse. Colon was harvested 72 hours after
irradiation with 1400 Gys. 1st strand cDNA was primed
with a Not I - oligo(dT) primer
[5'GTGTACGATCTGAGTGGAGGCGCGCCCTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors [AATTCGATCGTC], digested with Not I and cloned
into the Not I and Eco RI sites of the modified p1773
vector. Library constructed by Bob Barstead."
/db_xref="taxon:10090"
/clone="1137612"
/cdev_1lb="Barstead mouse irradiated colon MPLR37"
/dev_stage="8 weeks"
/lab_host="DH10B"

```


Query Match	52.6%	Score 113:	DB 12:	Length 488:
Best Local Similarity	42.9%	Pred. No. 3.28e-03:		
Matches	18:	Conservative	12:	Mismatches 12: Indels 0: Gaps 0:
Db	58	TGCAAGCATAGCATCCGTCTACACAAATAATGTCACATGAC	99	
Cp	50	TGNSMNMWYTCGATCNSMNMARNACNARATATGTCCKATTAAY	9	
RESULT	9			
LOCUS	M62849	426 bp	MRNA	EST
DEFINITION	Y283906.s1 Homo sapiens cDNA clone 289619 3'.			01-MAR-1996
ACCESSION	M62849			
NID	g1210678			
KEYWORDS	EST.			
SOURCE	human clone-289619 primer-m13 -40 forward library-Soares multiple sclerosis 2NBHSP vector-pt73p (pharmacia) with a modified polylinker V.TYPE: phagemid host-DH10B (ampicillin resistant) Bst1-Not I Rse12-Eco RI 46 year old male. 1st strand cDNA was primed with a Not I -Oligo(dT) primer (3'-TGTTACCATCTGAGTGGAGCGCCGATTTTCTTTTCTTTT-3'), double-stranded cDNA was size selected, ligated to Eco RI adapters (pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7 vector (pharmacia). Library went through one round of normalization to a Cot -5. library constructed by Bento Soares and M.Felima Bonaldo. RNA from 4 multiple sclerosis lesions from one patient was kindly provided by Dr. Kevin G. Becker (NINDS/NIH).			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostoma; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Channata; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Homidae; Homo. 1 (bases 1 to 426)			
AUTHORS	Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Ritzkin,L., Rohlfing,T., Soares,M., Tan,F., Treviskis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.			
TITLE	The WashU-Merck EST Project			
JOURNAL	Unpublished (1995)			
COMMENT	Contact: Wilson RK WashU-Merck EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.wustl.edu High quality sequence stops: 380 Source: IMAGE Consortium, LINTL This clone is available royalty-free through LINTL; contact the IMAGE Consortium (info@image.lnl.gov) for further information. Location/Qualifiers 1. 426 /organism="Homo sapiens" /clone="289619" <1..->426			
FEATURES	source			
BASE COUNT	MRNA 150 a 96 c 76 g 104 t			
ORIGIN				
Query Match	52.1%	Score 112:	DB 19:	Length 426:
Best Local Similarity	44.7%	Pred. No. 5.21e-03:		
Matches	17:	Conservative	11:	Mismatches 10: Indels 0: Gaps 0:
Db	234	ATACAAAAAATAATGTCCTATATAATACAGAA	271	
Cp	38	ATNSMNMARNACNARATATGTCCKATTAAYTCSWBA	1	
RESULT	10			
LOCUS	R37851	461 bp	MRNA	EST
				04-MAY-1995

```

DEFINITION y698bd02.s1 Homo sapiens cDNA clone 24106 3' similar to contains A1H repetitive element..
ACCESSION R37851
NID g/95307
KEYWORDS EST.
SOURCE human clone-24106 library-Souares infant brain INIB vector-Lafmid BA host-PHI0B (ampicillin resistant) primer--2lm3 Rsfelc-Not I Restr2-Hind III Whole brain from a 73 days post natal female. 1st strand cDNA was primed with a Not I - Oligo(CT) primer (5'-ACTGCAACAATTCGGCCGCAGCAGAATTGTGTGTGTTT TTTT TTTT TTTT 3'); double-stranded cDNA was ligated to Hind III adaptors (Pharmacia), digested with Not I and directionally cloned into the Not I and Hind III sites of the Lafmid BA vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo.
ORGANISM Homo sapiens
Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; HOMO.
REFERENCE 1 (bases 1 to 461)
AUTHORS Haller,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucab,T., Le,M., Lennon,G., Marra,M., Persons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.
TITLE The Wash-Merc EST Project
COMMENT Unpublished (1995)
DBS: G00-396-453
Contact: Wilson RK
Washu-Merc EST Project
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 347
Source: IMAGE Consortium, LMLU
This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES
source location/Qualifiers
1..461
/organism="Homo sapiens"
/clonetype="cDNA"
BASE COUNT 101 a 97 c 89 g 165 t 9 others
ORIGIN
Query Match 52.1%; Score 112; DB 5; Length 461;
Best Local Similarity 42.9%; Pred.No. 5, 2le=03;
Matches 15; Conservative 12; Mismatches 8; Indels 0; Gaps 0;
Db 16 TTTAGCTAATTTTTTGACACAGTAATTTGTCGTC 50
||||| :||: ||| |:||: ||| |::|
Qy 1 TTWTNGARTTYATMGMCARATYTTNTTNMWS 35
RESULT 11
LOCUS AQI73569 541 bp DNA GSS 09-SEP-1998
DEFINITION HSJ3202_A2.C03_T7 CHT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate-3202 Col-6 Row-E, genomic survey sequence.
ACCESSION AQI73569
NID g9370936
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; HOMO.
REFERENCE 1 (bases 1 to 541)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Furlong,J. and Shaker,R.
TITLE Construction of a Characterized Clone Resource for Genomic Sequencing
JOURNAL Unpublished (1998)
```


THIS PAGE BLANK (USPTO)

Molecular Biology
(TM)

Release 3.1a John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Msearch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Apr 16 14:25:17 1999; Msearch time 2.55 Seconds

Tabular output not generated.

Title: >US-09-011-797-4
Description: (1-8) from US09011797.dep
Perfect Score: 53
Sequence: 1 TLHONGNV 8

Scoring table: PAM 150
Gap 15

Searched: 131922 seqs, 16180660 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneeq2
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29

Statistics: Mean 14.066; Variance 35.673; scale 0.394

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	53	100.0	8 24	W25164	Human oploid receptor	2.48e+00
2	42	79.2	551 24	W18790	Corrected Bacillus la	7.03e+01
3	41	77.4	89 12	R68032	Mammalian MEK kinase	9.41e+01
4	41	77.4	224 9	R44533	Protein required for	9.41e+01
5	38	71.7	472 24	W26312	Rat STCH chaperone pr	2.33e+02
6	38	71.7	496 20	W01225	Aminopeptidase PRAP-4	2.33e+02
7	38	71.7	740 7	R38155	Acetobacter dlyanula	2.23e+02
8	38	71.7	941 25	W31559	Xenopus frog protein	2.23e+02
9	38	71.7	1319 8	R45002	Cellulose synthase op	2.23e+02
10	37	69.8	15 16	R92307	Tryptic peptide from	2.96e+02
11	37	69.8	532 9	R47237	Wild-type Feline Hep	2.96e+02
12	37	69.8	532 13	R36588	Feline herpesvirus g1	2.96e+02
13	37	69.8	716 13	R71616	Thermotoga pullulan	2.96e+02
14	37	69.8	816 27	W37501	Human nel-related pro	2.96e+02
15	37	69.8	839 16	R75109	Glycosyl-phosphatidyl	2.96e+02
16	37	69.8	839 4	R22275	Bovine liver GPI-PD.	2.96e+02
17	37	69.8	1129 13	R70830	Murine JAK2 kinase.	2.96e+02
18	36	67.9	243 25	W15799	Adherence factor 10AR	3.92e+02

19	36	67.9	286 1	P82590	Polypeptide with glyc	3.92e+02
20	36	67.9 <td>309 26<td>W21741<td>Neisseria meningitidis</td><th>3.92e+02</th></td></td>	309 26 <td>W21741<td>Neisseria meningitidis</td><th>3.92e+02</th></td>	W21741 <td>Neisseria meningitidis</td> <th>3.92e+02</th>	Neisseria meningitidis	3.92e+02
21	36	67.9 <td>309 13<td>R70762<td>Meningococcal group B</td><th>3.92e+02</th></td></td>	309 13 <td>R70762<td>Meningococcal group B</td><th>3.92e+02</th></td>	R70762 <td>Meningococcal group B</td> <th>3.92e+02</th>	Meningococcal group B	3.92e+02
22	36	67.9 <td>313 26<td>W21744<td>Neisseria meningitidis</td><th>3.92e+02</th></td></td>	313 26 <td>W21744<td>Neisseria meningitidis</td><th>3.92e+02</th></td>	W21744 <td>Neisseria meningitidis</td> <th>3.92e+02</th>	Neisseria meningitidis	3.92e+02
23	36	67.9 <td>316 16<td>R75111<td>Glycosyl-phosphatidyl</td><th>3.92e+02</th></td></td>	316 16 <td>R75111<td>Glycosyl-phosphatidyl</td><th>3.92e+02</th></td>	R75111 <td>Glycosyl-phosphatidyl</td> <th>3.92e+02</th>	Glycosyl-phosphatidyl	3.92e+02
24	36	67.9 <td>316 4</td> <td>R22277<td>Human pancreatic GPI-</td><th>3.92e+02</th></td>	316 4	R22277 <td>Human pancreatic GPI-</td> <th>3.92e+02</th>	Human pancreatic GPI-	3.92e+02
25	36	67.9 <td>350 17<td>R92754<td>Human growth differen</td><th>3.92e+02</th></td></td>	350 17 <td>R92754<td>Human growth differen</td><th>3.92e+02</th></td>	R92754 <td>Human growth differen</td> <th>3.92e+02</th>	Human growth differen	3.92e+02
26	36	67.9 <td>445 24<td>W27163<td>Human TRAF inhibitor</td><th>3.92e+02</th></td></td>	445 24 <td>W27163<td>Human TRAF inhibitor</td><th>3.92e+02</th></td>	W27163 <td>Human TRAF inhibitor</td> <th>3.92e+02</th>	Human TRAF inhibitor	3.92e+02
27	36	67.9 <td>441 26<td>W34856<td>Human tau protein.</td><th>3.92e+02</th></td></td>	441 26 <td>W34856<td>Human tau protein.</td><th>3.92e+02</th></td>	W34856 <td>Human tau protein.</td> <th>3.92e+02</th>	Human tau protein.	3.92e+02
28	36	67.9 <td>583 2</td> <td>R07999<td>Asparagine synthetase</td><th>3.92e+02</th></td>	583 2	R07999 <td>Asparagine synthetase</td> <th>3.92e+02</th>	Asparagine synthetase	3.92e+02
29	36	67.9 <td>586 2</td> <td>R07998<td>Asparagine synthetase</td><th>3.92e+02</th></td>	586 2	R07998 <td>Asparagine synthetase</td> <th>3.92e+02</th>	Asparagine synthetase	3.92e+02
30	36	67.9 <td>841 4</td> <td>R22276<td>Human liver GPI-PD.</td><th>3.92e+02</th></td>	841 4	R22276 <td>Human liver GPI-PD.</td> <th>3.92e+02</th>	Human liver GPI-PD.	3.92e+02
31	36	67.9 <td>841 16<td>R75110<td>Glycosyl-phosphatidyl</td><th>3.92e+02</th></td></td>	841 16 <td>R75110<td>Glycosyl-phosphatidyl</td><th>3.92e+02</th></td>	R75110 <td>Glycosyl-phosphatidyl</td> <th>3.92e+02</th>	Glycosyl-phosphatidyl	3.92e+02
32	36	67.9 <td>1539 21<td>W13905<td>Human SKcy protein.</td><th>3.92e+02</th></td></td>	1539 21 <td>W13905<td>Human SKcy protein.</td><th>3.92e+02</th></td>	W13905 <td>Human SKcy protein.</td> <th>3.92e+02</th>	Human SKcy protein.	3.92e+02
33	36	67.9 <td>4987 3</td> <td>R10834<td>Ranodin receptor.</td><th>3.92e+02</th></td>	4987 3	R10834 <td>Ranodin receptor.</td> <th>3.92e+02</th>	Ranodin receptor.	3.92e+02
34	36	67.9 <td>5035 5</td> <td>R25450<td>MH mutant porcine rya</td><th>3.92e+02</th></td>	5035 5	R25450 <td>MH mutant porcine rya</td> <th>3.92e+02</th>	MH mutant porcine rya	3.92e+02
35	36	67.9 <td>5072 2</td> <td>R11510<td>Ryanodine receptor de</td><th>3.92e+02</th></td>	5072 2	R11510 <td>Ryanodine receptor de</td> <th>3.92e+02</th>	Ryanodine receptor de	3.92e+02
36	36	66.0 <td>50 26</td> <td>P30660<td>Enkephalin-like Pepti</td><th>5.16e+02</th></td>	50 26	P30660 <td>Enkephalin-like Pepti</td> <th>5.16e+02</th>	Enkephalin-like Pepti	5.16e+02
37	35	66.0 <td>59 24</td> <td>W22623<td>Type II topoisomerase</td><th>5.16e+02</th></td>	59 24	W22623 <td>Type II topoisomerase</td> <th>5.16e+02</th>	Type II topoisomerase	5.16e+02
38	35	66.0 <td>282 17<td>R95450<td>Bacteriophage resistin</td><th>5.16e+02</th></td></td>	282 17 <td>R95450<td>Bacteriophage resistin</td><th>5.16e+02</th></td>	R95450 <td>Bacteriophage resistin</td> <th>5.16e+02</th>	Bacteriophage resistin	5.16e+02
39	35	66.0 <td>351 28<td>W34209<td>Streptomyces acyltran</td><th>5.16e+02</th></td></td>	351 28 <td>W34209<td>Streptomyces acyltran</td><th>5.16e+02</th></td>	W34209 <td>Streptomyces acyltran</td> <th>5.16e+02</th>	Streptomyces acyltran	5.16e+02
40	35	66.0 <td>398 4</td> <td>R22322<td>Marek Disease Virus U</td><th>5.16e+02</th></td>	398 4	R22322 <td>Marek Disease Virus U</td> <th>5.16e+02</th>	Marek Disease Virus U	5.16e+02
41	35	66.0 <td>512 22</td> <td>W15274</td> <td>Salmonella secreted p</td> <th>5.16e+02</th>	512 22	W15274	Salmonella secreted p	5.16e+02
42	35	66.0 <td>541 26<td>W31507<td>Human receptor protei</td><th>5.16e+02</th></td></td>	541 26 <td>W31507<td>Human receptor protei</td><th>5.16e+02</th></td>	W31507 <td>Human receptor protei</td> <th>5.16e+02</th>	Human receptor protei	5.16e+02
43	35	66.0 <td>833 6<td>R28960<td>Delta DII.</td><th>5.16e+02</th></td></td>	833 6 <td>R28960<td>Delta DII.</td><th>5.16e+02</th></td>	R28960 <td>Delta DII.</td> <th>5.16e+02</th>	Delta DII.	5.16e+02
44	35	66.0 <td>903 28<td>W37391<td>Human additional sex</td><th>5.16e+02</th></td></td>	903 28 <td>W37391<td>Human additional sex</td><th>5.16e+02</th></td>	W37391 <td>Human additional sex</td> <th>5.16e+02</th>	Human additional sex	5.16e+02
45	35	66.0 <td>903 28<td>W37389<td>Human additional sex</td><th>5.16e+02</th></td></td>	903 28 <td>W37389<td>Human additional sex</td><th>5.16e+02</th></td>	W37389 <td>Human additional sex</td> <th>5.16e+02</th>	Human additional sex	5.16e+02

ALIGNMENTS

RESULT 1
ID W25164 standard; peptide; 8 AA.
AC W25164;

DT 08-DEC-1997 (first entry)
DE Human oploid receptor-like 1 receptor ligand, nociceptin, fragment.
KW Oploid receptor-like 1; ORL1; ligand; pronociceptive; stress;
KW hyperalgesia; locomotor activity disease; anxiety; neuroendocrine;
KW memory; attention; sensory perception; learning; homeostasis;
KW hypoalgesia; nociceptin; endorphin; dynorphin A.
OS Synthetic.
PN W09707208-A1.
PD 27-FEB-1997.
PF 14-AUG-1996; BE0087.
PR 15-AUG-1995; US-002368.

PA (ULBR) UNIV LIBRE BRUXELLES.
PI Meunier J, Mollereau C, Parmentier M, Vassart G;
DR WPI, 97-165292/15.
DR N-PSDB; T79888.

PT Novel ligand for the oploid receptor-like receptor, nociceptin -
PT has pro-nociceptive properties, useful for treating or preventing
PT diseases related to e.g. stress, hyperalgesia, locomotor activity,
PT etc.

PS Claim 8, page 27; 48pp; English.
CC W25162-W25164 are fragments of a ligand to human oploid receptor-like 1
CC (ORL1) receptor, designated nociceptin, which resembles the endorphin
CC dynorphin A. Inhibitors of nociceptin or its DNA, e.g. antisense
CC sequences and antibodies can be used as new types of drugs in the
CC control of various behaviours or functions. The inhibitors can be used
CC to prevent or treat a disease related to hyperalgesia, neuroendocrine
CC secretion, stress, locomotor activity, anxiety, instinctive behaviour,
CC learning and memory, homeostasis, hyperalgesia, hypoalgesia and/or
CC sensory perception.
SQ Sequence 8 AA:

Query Match 100.0%; Score 53; DB 24; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.48e+00;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 tlhngnv 8
| | | | |
QY 1 TLHONGNV 8

RESULT 2
ID W18790 standard; protein; 551 AA.
AC W18790.
DT 18-NOV-1997 (first entry)
DE Corrected Bacillus laevis (NCIMB 40250) endoglucanase Endo 3A.
KW Endoglucanase; Endo 3A; formation; localised; variation;
KW colour density; surface; dye; fabric; family 5; cellulose;
KW hydrolystase; p-nitrophenyl-beta-1,4-cellobioside; stone wash;
KW blue jeans; back staining.
OS Bacillus laevis.
PN W09709410-A1.
PD 13-MAR-1997.
PE 03-SEP-1996; DK0364.
PR 08-SEP-1995; DK-000993.
PA (NOVO) NOVO-NORDISK AS.
PI Fitch M, Onishi M, Schulein M, Toft AH;
DR WPI: 97-192888/17.
PT localised variation of colour density in the surface of a dyed
PT cellulose fabric - uses cellulase compsn. able to hydrolyse
PS p-nitrophenyl-beta-1,4-cellobioside
CC Disclosure; Pages 15-17; 23pp; English.
CC The present sequence is the corrected version of the incorrect
CC Bacillus laevis (NCIMB 40250) endoglucanase Endo 3A described in
CC W09110732. Endo 3 can be used in novel method of forming localised
CC colour density variation on the surface of a dyed cellulose
CC fabric. The method comprises agitating the fabric in an aqueous
CC medium (pH 6.5 to 9.0) containing a family 5 cellulose,
CC e.g. the present sequence, which can hydrolyse p-nitrophenyl-beta
CC -1,4-cellobioside, or a family 7 cellulase, and a mechanical
CC abrading agent or cellulose having abrading activity. Each
CC cellulase displays 30 % or more of its maximum activity at pH 7.
CC The process is useful to provide a stone washed look to blue jeans
CC without back staining.
SQ Sequence 551 AA;

Query Match 79.2%; Score 42; DB 24; Length 551;
Best Local Similarity 75.0%; Pred. No. 7.03e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 538 lthngnvl 545
QY 1 lthngnvn 8

RESULT 3
ID R66032 standard; protein; 89 AA.
AC R66032.
DT 28-JUN-1995 (first entry)
DE Mammalian MEK kinase (MEKK 4) catalytic kinase domain.
KW MEK kinase; MEKK 4; mitogen-activated protein kinase regulator;
KW MAPK; cell atrophy inhibition; Parkinson's; Alzheimer's; cancer;
KW autoimmune diseases; allergies; wound healing; oncogenes;
KW tumour agents; neurotropic growth factor.
OS Mus musculus.
PN W09424159-A.
PD 27-OCT-1994.
PE 15-APR-1994; U04178.
PR 15-APR-1993; US-049254.
PA (NAJE-) NAT JEWISH CENT IMMUNOLOGY & RESPIRATORY.
PI Johnson GL;
DR WPI: 94-357747/44.
PT New MEK kinase protein and related antibodies and nucleic acid
PT regulator of mitogen activated protein kinase, useful
PT therapeutically to inhibit cell atrophy, to screen for oncogenes
PT etc.
PS Claim 6; Page 15; 84pp; English.
CC 079925 encodes R66032 the mammalian MEK kinase (MEKK 1), other
CC unique mammalian MEK kinases identified by PCR are described in
CC R66030 (MEKK 2), R66031 (MEKK 3) and R66032 (MEKK 4). MEKK is an
CC activator, independent of Raf protein, of mitogen-activated protein
CC kinases (MAPK). Inactivation of MEKK can be used in the treatment
CC of some cancers, autoimmune diseases and allergies, while

CC stimulation can promote wound healing. MEKK can also be used to
CC alleviate cellular atrophy in Parkinson's or Alzheimer's by acting
CC as a neurotropic growth factor, and to screen for oncogenes and
CC tumour agents.
SQ Sequence 89 AA;

Query Match 77.4%; Score 41; DB 12; Length 89;
Best Local Similarity 71.4%; Pred. No. 9.41e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 83 lthngnv 89
QY 2 lthngnvn 8

RESULT 4
ID R44533 standard; protein; 224 AA.
AC R44533.
DT 24-JUN-1994 (first entry)
DE Protein required for surfactin synthesis.
KW Surfactin; Bacillus subtilis; antibiotic; fungicide; anticoagulant;
KW thrombosis; myocardial infarction; embolism.
OS Bacillus subtilis.
PN EP-576050-A.
PD 29-DEC-1993.
PE 23-APR-1993; 201177.
PR 24-APR-1992; IT-MI0976.
PA (ENIE) ENRICEERCE SPA.
PI Cosmina P, De Ferra F, Grandi G, Perego M, Rodriguez F;
DR WPI: 94-001231/01.
DR N-PEDB: Q39499.
PT DNA from Bacillus subtilis - encodes surfactin synthesis protein
PS Claim 1; Figure 4; 24pp; English.
CC The protein is required for surfactin synthesis. The coding
CC sequence (sfp gene, Q39499) is useful for the in vitro and in vivo
CC production of surfactin and analogues having anti-cholesterase,
CC functional and antibiotic properties. It can also be used as an
CC anticoagulant for thrombosis and for prophylaxis of myocardial
CC infarcts and pulmonary emboli. This protein was isolated from a
CC producing strain of microorganism.
SQ Sequence 224 AA;

Query Match 77.4%; Score 41; DB 9; Length 224;
Best Local Similarity 71.4%; Pred. No. 9.41e+01;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 172 lthngnv 178
QY 2 lthngnvn 8

RESULT 5
ID W26312 standard; protein; 472 AA.
AC W26312.
DT 17-NOV-1997 (first entry)
DE Rat STCH chaperone protein.
KW STCH; stress 70 protein; chaperone protein; protein folding; rat.
OS Rattus sp.
PN US5646249-A.
PD 28-FEB-1994; 203905.
PE 28-FEB-1994; US-203905.
PR 28-FEB-1994; US-203905.
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PI Kaye FU, Oeterson GA;
DR WPI: 97-362996/33.
DR N-PSDB: T84446.
PT Recombinant Stress 70 Chaperone family STCH proteins - useful as
PT chaperone proteins for facilitating protein folding
PS Claim 1; Column 47-50; 30pp; English.
CC Rat STCH (W26312) is a novel member of the stress 70 chaperone
CC protein family and is characterised by a truncated C-terminal
CC peptide binding domain, peptide-independent core ATPase activity
CC and a hydrophobic leader peptide. STCH over-expression suppresses

CC oncogene transformation. STCH expressed in cells after induction
 CC with calcium ionophore A23187 (but not following heat shock) and
 CC localises within the microsome lumen of a cell. It shows over 90%
 CC homology to human STCH (W26311). Recombinant STCH can be used to
 CC facilitate protein folding, e.g. during in vitro translation and
 CC in dialysis following urea extraction, or for x-ray crystallography.
 CC The leader peptide can be used to express a desired protein in the
 CC endoplasmic reticulum. Antibodies raised against STCH can be used
 CC for STCH detection and cell localisation.
 SQ Sequence 472 AA;

Query Match 71.7%; Score 38; DB 24; Length 472;
 Best Local Similarity 62.5%; Pred. No. 2.23e+02;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 301 tlhgsagv 308
 ||||:|
 OY 1 TLHONGNV 8

RESULT 6
 ID W01225 standard; Protein; 496 AA.
 AC W01225;
 DT 16-MAR-1997 (first entry)
 DE Aminopeptidase PfAP-496.
 KW Flea; midgut; aminopeptidase; PfAP-496; vaccine;
 KM domestic animal; infestation; insecticide; protease-inhibitor;
 KM controlled release formulation; synergist.
 OS Siphonaptera sp.
 FH Key Location/Qualifiers
 FT Peptide 1..15
 FT Peptide 1..151
 FT Peptide /note- "Signal peptide"
 FT Peptide /note- "nFAP-151 (W01224, claim 71)."

PN W09611706-A1.
 PD 25-APR-1996
 PF 18-OCT-1995
 PR 18-OCT-1994; US-326773.
 PR 07-JUN-1995; US-485455.
 PR 07-JUN-1995; US-482130.
 PR 07-JUN-1995; US-485443.
 PR 07-JUN-1995; US-484211.
 PA (PARA-) PARAVAX INC.
 PA (HESK-) HESKA CORP.
 PI Arsten A, Dale B, Frank GR, Grieve RB, Heath A;
 PI Hunter SM, Rushlow KE, Stiegler GL, Yamana M;
 DR WPI: 96-221762/22.
 DR N-PSDB: T40861.

PT DNA encoding Flea serine protease and aminopeptidase - useful in
 PT vaccines to protect animals from flea infestation.
 PS Claim 71: Page 199-200; 241pp; English.
 CC This sequence (PfAP-496) represents a flea midgut aminopeptidase
 CC encoded by the nFAP-1580 gene, isolated from a flea cDNA library by
 CC homology with a cattle lens leucine-aminopeptidase conserved region.
 CC The sequence contains a signal peptide, and also contains N-terminal
 CC sequence PfAP-151 (W01224). The peptide may be used in a vaccine for
 CC protection of domestic animals from flea infestation, or in isolation
 CC of protease-inhibitors, which may be used in controlled release
 CC formulations to reduce the flea burden on and around the animal. The
 CC inhibitors may be included in insecticidal compositions to increase
 CC efficiency of other active agents, by reducing proteolytic activity in
 CC the flea midgut.

SQ Sequence 496 AA;

Query Match 71.7%; Score 38; DB 20; Length 496;
 Best Local Similarity 62.5%; Pred. No. 2.23e+02;

Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 394 tlhagrv 401
 ||||:|
 OY 1 TLHONGNV 8

RESULT 7
 ID R38155 standard; Protein; 740 AA.
 AC R38155;
 DT 13-OCT-1993 (first entry)

DE Acetobacter diacylate; diacylate phosphodiesterase PDEA3.
 KW Cyclic diacylate; diacylate phosphodiesterase;
 KW diacylate cyclase; cellulose production; cdg3 operon.
 OS Acetobacter xylinum.
 FH Key Location/Qualifiers
 FT Peptide 119
 FT Peptide /note- "Val deduced from GT"
 FT Peptide /note- "Trp deduced from GG"
 FT Peptide /note- "Trp deduced from GG"
 FT Peptide /note- "Ala deduced from CC"

PN W09311244-A.
 PD 10-JUN-1993
 PF 14-OCT-1992; U08756.
 PR 29-NOV-1991; US-800218.
 PA (WEYE) WEYERHAEUSER CO.
 PI Ben-Bassat A, Ben-Ziman M, Calhoun RD, Gelfand DH;
 PI Tal R, Wong HC;
 DR WPI: 93-197062/24.
 DR N-PSDB: Q43662.

PT Polynucleotide sequence from Acetobacter cdg operon - encodes
 PT cyclic diacylate mono:phosphate degradation enzymes e.g.
 PT 3-phosphodiesterase isozyme
 PS Claim 5: Page 86-88; 98pp; English.
 CC The amino acid sequence of protein PDEA3 was deduced from the 1st.
 CC open reading frame of the cdg3 operon. The protein is a diacylate
 CC phosphodiesterase A, i.e. it enzymatically cleaves a single
 CC phosphodiester bond in c-di-GMP to yield the linear dimer pGpG.
 CC See also R38156.
 SQ Sequence 740 AA;

Query Match 71.7%; Score 38; DB 7; Length 740;
 Best Local Similarity 71.4%; Pred. No. 2.23e+02;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 60 tlhagrv 66
 ||||:|
 OY 1 TLHONGNV 7

RESULT 8
 ID W31558 standard; Protein; 941 AA.
 AC W31558;
 DT 16-FEB-1998 (first entry)

DE Xenopus frog protein "chordin"
 KW Xenopus protein; chordin; dorsal tissue; neural tissue; vertebrate;
 KW endodermal differentiation; treatment; neurodegenerative disease;
 KW nerve cell; transforming growth factor; TGF; secreted protein.
 OS Xenopus laevis.

FH Key Location/Qualifiers
 FT Peptide 1..19
 FT Peptide /note- "hydrophobic signal peptide"
 FT Protein /note- "putative secreted protein chordin"

PN U55679783-A.
 PD 21-OCT-1997
 PF 22-NOV-1994; US-343760.
 PR 22-NOV-1994; US-343760.
 PA (REGC) UNIV CALIFORNIA.
 PI De Robertis EM, Sasal Y;
 DR WPI: 97-525754/48.
 DR N-PSDB: T93499.

PT DNA encoding Xenopus frog protein - that induces dorsal and neural
 PT development and endodermal differentiation in vertebrates
 PS Claim 2: Columns 23-28; 27pp; English.
 CC This is the Xenopus protein "chordin". This functional recombinant
 CC protein chordin has a defined sequence of 941 amino acids and can induce
 CC dorsal and neural development and endodermal differentiation in
 CC vertebrates. The presence of a hydrophobic signal sequence, four possible

CC N-glycosylation sites and conserved Cys-rich repeat regions suggest that
 CC chordin is a secreted protein. The encoding DNA sequence can be
 CC operationally linked with an expression vector, to form a construct and a
 CC transfectant can be obtained by introducing the construct into a host.
 CC Chordin may be useful as a component of culture media for culturing cells
 CC such as nerve or muscle cells, for treating neurodegenerative diseases
 CC and damaged nerve cells.
 SO Sequence 941 AA.

Query Match 71.7%; Score 38; DB 25; Length 941;
 Best Local Similarity 62.5%; Pred. No. 2.23e+02;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 429 tlnhgntl 436
 QY 1 TLHONGNV 8

RESULT 9
 ID R45002 standard; Protein: 1319 AA.

AC R45002:
 DT 13-YUN-1994 (first entry)
 DE Cellulose synthase operon, gene C product.
 KW Bacterial; cellulose synthase; operon; gene A; gene B; gene C;
 KW gene D; transcription vector; recombinant microorganism;
 KW cellulose synthesis.
 OS Acetobacter xylinum.
 PN US568274-A.
 PD 07-DEC-1993.
 PF 12-APR-1989; 337194.
 PR 04-APR-1990; WO-001811.
 PR 09-APR-1990; IL-094053.
 PR 10-APR-1990; CA-014264.
 PR 11-APR-1990; IE-001317.
 PR 12-APR-1990; NZ-233312.
 PA (CERTU) CERTUS CORP.
 PI Ben-Bassat A, Ben-Ziman M, Calhoun RD, Fear AL, Gelfand DH;
 PI Meade JH, Tai R, Wong H;
 DR WPI: 93-404004/50.
 DR N-PSDB: 053522.
 PT Contiguous nucleic acid sequences - encoding bacterial cellulose
 PT synthase
 PS Claim 52; Fig 1 and Columns 53-72; 79pp; English.
 CC The sequences given in R45000-03 represent the proteins encoded by
 CC the bacterial cellulose synthase operon. The sequence given in
 CC R45004 is the beginning of an open reading frame overlapping the end
 CC of this operon. The bacterial cellulose synthase operon contains
 CC four genes, genes A-D. The operon sequence may be used in a
 CC transcription vector for the expression of the cellulose synthase
 CC operon to increase cellulose production in a recombinant microorganism.
 CC This system may be used as an important tool for exploring mechanisms
 CC of cellulose synthesis and for enhancing production of cellulose.
 SO Sequence 1319 AA;

Query Match 71.7%; Score 38; DB 8; Length 1319;
 Best Local Similarity 50.0%; Pred. No. 2.23e+02;
 Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 518 alggqgdv 525
 QY 1 TLHONGNV 8

RESULT 10
 ID R92307 standard; Peptide: 15 AA.

AC R92307:
 DT 17-MAY-1996 (first entry)
 DE Tryptic peptide from bovine serum phospholipase-D.
 KW Tryptic peptide; glycosyl-phosphatidylinositol; phospholipase-D;
 KW bovine; serum; probe; hybridisation; liver; signal peptide;
 KW recombinant protein; anchor; vector; cloning; protein secretion;
 KW fusion protein; purification.
 SO Bos taurus.

PN US5418147-A.
 PD 23-MAY-1995.
 PR 27-SEP-1990; 588896.
 PR 27-SEP-1990; US-588896.
 PR 31-MAR-1992; US-860825.
 PA (HOFF) HOFFMANN LA ROCHE INC.
 PI Huang K, Kochan JP, Li SH, Pan YE, Scallon BJ;
 PI Tsang TCH;
 DR WPI: 95-199737/26.
 PT New glycosyl-phosphatidyl:inositol-specific phospholipase D gene -
 PT used partic. for producing secreted recombinant proteins from
 PT GPI-anchored proteins expressed by cells
 PS Example 1; Column 19-20; 54pp; English.
 CC The sequence represents a fraction-T34 tryptic peptide derived from
 CC bovine serum glycosyl-phosphatidylinositol (GPI)-specific
 CC phospholipase-D. The sequence, along with those of peptides
 CC R7514-26, is used in construction of probes for isolation of the
 CC corresponding full-length bovine liver gene (encoding R75109). The
 CC resulting gene may be connected to a heterologous gene linked at the
 CC 3'-end to a C-terminal signal peptide (e.g. R75129 or R75130) gene to
 CC form a GPI-anchored molecule, where the C-terminal peptide directs
 CC attachment of a GPI anchor to the heterologous protein. The resulting
 CC construct may be inserted in a vector for expression in a host cell.
 CC The enzyme selectively hydrolyses the inositol-phosphate linkage of
 CC GPI-anchored proteins, GPI lipids and related molecules, resulting in
 CC direct secretion into the medium and simple recombinant protein
 CC purification.
 SO Sequence 15 AA;

Query Match 69.8%; Score 37; DB 16; Length 15;
 Best Local Similarity 71.4%; Pred. No. 2.96e+02;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 7 tlnhgns 13
 QY 1 TLHONGNV 7

RESULT 11
 ID R47237 standard; Protein: 532 AA.
 AC R47237:
 DT 06-SEP-1994 (first entry)
 DE Wild-type Feline Herpes Virus ORF-2-encoded protein.
 KW Feline herpes virus; FHV; genome; mutant; vaccine; ORF-2;
 KW feline viral rhinotracheitis; herpesviridae family;
 KW herpes virus A subfamily.
 OS Feline Herpes Virus.
 PN WO9403621-A.
 PD 17-FEB-1994.
 PF 23-JUL-1993; E01971.
 PR 30-JUL-1992; EP-202365.
 PA (ALKO) AKZO NV.
 PI Sondermeijer PJA, Willemsse MJ;
 PI WPI: 94-065709/08.
 DR N-PSDB: 056188.
 PT Feline herpes virus mutant comprising a heterologous gene
 PT inserted in the virus genome - for vaccination against feline
 PT herpes virus and other feline pathogens.
 PS Claim 2; Page 39-40; 55pp; English.
 CC Mutant versions of the wild-type feline herpes virus genome (056188)
 CC are claimed. Esp. the FHV mutant is one which fails to produce one
 CC or more antigenic or functional polypeptides. The mutant may contain
 CC a heterologous nucleic acid sequence inserted within one of the 6
 CC open reading frames. The attenuated FHV mutants can be used to prepare
 CC vaccines against feline viral rhinotracheitis and, where the
 CC heterologous insert encodes an antigen of a feline pathogen,
 CC against other infectious diseases of feline.
 SO Sequence 532 AA;

Query Match 69.8%; Score 37; DB 9; Length 532;
 Best Local Similarity 57.1%; Pred. No. 2.96e+02;
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 147 tlhngd 153
 |||:|:
 QY 1 TLHQNGN 7

RESULT 12
 ID R96588 standard; Protein: 532 AA.

AC R96588;
 DT 17-DEC-1996 (first entry)
 DE feline herpesvirus glycoprotein E.
 KW feline herpesvirus; FHV; deletion; insertion; glycoprotein E;
 KW gp E; unique short region; attenuated virus; vaccine; protection;
 KW cat; vector; delivery; antigen.
 OS feline herpesvirus.
 PN MO9613575-A1.
 PD 09-MAY-1996.
 PF 26-OCT-1995; U13975.
 PR 26-OCT-1994; US-329883.
 PA (SYTR) SYNTRO CORP.
 PI Cochran MD. McDonnell MW;
 DR WPI: 96-239489/24.
 DR N-PSDB; T28190.
 PT Recombinant feline herpes virus attenuated by alteration of the gp E gene - and related homology vectors, useful in vaccines, also as vectors for delivering antigens or therapeutic agents to mammals and birds
 PS Example 1; Pages 76-78; 162pp; English.
 CC A novel feline herpesvirus (FHV) contains a deletion or insertion in the glycoprotein E (gp E) gene, in the unique short region of the viral genome. Altering the gp E gene, so that it no longer produces functional gp E.1.e. the present sequence, attenuates the virus. The attenuated virus is useful in vaccines to protect cats against FHV (and opt. against other diseases if appropriate DNA is inserted), and as a vector for delivering vaccinating antigens or therapeutic agents (e.g. antisense mols., ribozymes, interferon inducers, hormones, lymphokines, etc.) to other mammals (including humans) or birds.
 CC humans) or birds.
 CC Sequence 532 AA;
 SQ

Query Match 69.8%; Score 37; DB 19; Length 532;
 Best Local Similarity 57.1%; Pred. No. 2.96e+02;
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 Db 147 tlhngd 153
 |||:|:
 QY 1 TLHQNGN 7

RESULT 13
 ID R71616 standard; Protein: 716 AA.
 AC R71616;
 DT 10-NOV-1995 (first entry)
 DE Thermostable pululanase.
 KW pululanase; thermostable.
 OS Thermus sp.
 PN J07023783-A.
 PD 27-JAN-1995.
 PF 15-JUL-1991; 200089.
 PR 15-JUL-1991; JP-200089.
 PA (NISO) NIPPON SHOKUHIN KAKO KK.
 PA (OPPP) OP CORP.
 PA (RIKA) RIKAGAKU KENKYUSHO.
 PA (SHKO) SHINGIUTSU JIGYODAN.
 DR WPI: 95-100945/14.
 DR N-PSDB; O85999.
 PT A thermostable pululanase gene - for the production of recombinant pululanase free from amylase activity
 PS Claim 1; Fig 2-4; 9pp; Japanese.
 CC The new thermostable pululanase gene (O85999) can be produced by recombinant DNA techniques. Esp. the pululanase prepd. has no amylase activity.
 CC Sequence 716 AA;
 SQ

Query Match 69.8%; Score 37; DB 13; Length 716;
 Best Local Similarity 83.3%; Pred. No. 2.96e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Db 328 tlhng 333
 |||:|:
 QY 1 TLHQNG 6

RESULT 14
 ID W37501 standard; Protein: 816 AA.

AC W37501;
 DT 20-APR-1998 (first entry)
 DE Human nel-related protein type 2.
 KW Human; foetal brain cDNA library; GDP dissociation stimulating protein;
 KW brain specific nucleosome assembly protein; diagnosis; therapy;
 KW skeletal muscle specific ubiquitin conjugating enzyme; RMP-2; NPK;
 KW nel-related protein type 1; nel-related type 2; hereditary disease;
 KW cancer.
 OS Homo sapiens.
 PN EP-796913-A2.
 PD 24-SEP-1997.
 PF 19-MAR-1997; 104842.
 PR 05-MAR-1997; JP-069163.
 PR 19-MAR-1996; JP-063410.
 PA (SARA) OTSUKA PHARM CO LTD.
 PI Fujiwara T, Horie M, Watanabe T;
 DR WPI: 97-459830/43.
 DR N-PSDB; V01882, V01883.
 PT Novel human genes, e.g. brain-specific nucleosome assembly protein - useful for diagnosis or therapy of hereditary disease and cancer
 PS Claim 22; Page 104-107; 123pp; English.
 CC The present sequence represents a nel-related protein type 2 isolated from a human foetal brain cDNA library. The nucleotide or amino acid sequences are useful for in-vitro diagnosis of hereditary diseases and cancer and for preparation of pharmaceuticals.
 CC Sequence 816 AA;
 SQ

Query Match 69.8%; Score 37; DB 27; Length 816;
 Best Local Similarity 83.3%; Pred. No. 2.96e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Db 701 lhng 706
 |||:|:
 QY 2 TLHQNGN 7

RESULT 15
 ID R75109 standard; Protein: 839 AA.
 AC R75109;
 DT 15-MAY-1996 (first entry)
 DE Glycosyl-phosphatidylinositol-specific phospholipase-D.
 KW Glycosyl-phosphatidylinositol; phospholipase-D; bovine; liver;
 KW signal peptide; recombinant protein; metal ion binding domain;
 KW tryptic peptide; N-glycosylation; anchor; vector; cloning;
 KW fusion protein; protein secretion; purification.
 OS Bos taurus.
 FH Key
 FT peptide 1..23 Location/Qualifiers
 FT peptide /note= "Signal peptide"
 FT peptide /note= "N-terminal peptide homologous to R75112"
 FT peptide /note= "36..50
 FT peptide /note= "Tryptic peptide R75119"
 FT peptide /note= "56..70
 FT peptide /note= "Tryptic peptide R75120"
 FT peptide /note= "56..77
 FT peptide /note= "Tryptic peptide R75121"
 FT peptide /note= "126..138
 FT peptide /note= "Tryptic peptide R75116"
 FT peptide /note= "236..261
 FT peptide /note= "Tryptic peptide R75113"
 FT peptide /note= "349..358"


```

FT domain /note="Tryptic peptide R75114"
FT 379..402
FT /note="Putative metal ion binding domain"
FT peptide 380..408
FT /note="Tryptic peptide R75116"
FT domain 449..471
FT /note="Putative metal ion binding domain"
FT peptide 449..477
FT /note="Tryptic peptide Q75118"
FT domain 512..534
FT /note="Putative metal ion binding domain"
FT peptide 623..636
FT /note="Tryptic peptide R75115"
FT 678..692
FT /note="Tryptic peptide R75122"
FT domain 717..739
FT /note="Putative metal ion binding domain"
FT peptide 786..798
FT /note="Tryptic peptide R75123"
FT 799..807
FT /note="Tryptic peptide R75124"

FT US5418147-A.
FT 23-MAY-1995.
FT 27-SEP-1990: 588896.
FT 27-SEP-1990: US-588896.
FT 31-MAR-1992: US-860825.
FT (HOF) HOFFMANN LA ROCHE INC.
FT Huang K, Kochan JP, Li SH, Pan YE, Scalton BJ;
FT Tsang TCH;
FT WPI: 95-199737/26.
FT N-PSDB: 090582.
FT New glycosyl-phosphatidylinositol-specific phospholipase D gene -
FT used partic. for producing secreted recombinant proteins from
FT GPI-anchored proteins expressed by cells
FT Claim 1; Fig 5; 54pp; English.
FT The sequence represents a bovine liver glycosyl-phosphatidylinositol
FT (GPI)-specific phospholipase-D. The sequence contains an
FT N-terminal signal peptide and 4 regions showing sequence similarity
FT to metal ion binding domains of integrin-a subunits, with a core
FT sequence (R75127) matching the EF-hand consensus motif
FT characteristic of calcium and magnesium binding proteins. Tryptic
FT peptides R75113-26 have been derived from the native purified protein.
FT The sequence contains 8 putative N-glycosylation sites. A gene
FT encoding the enzyme may be connected to a heterologous gene linked at
FT the 3'-end to a C-terminal signal peptide (e.g. R75129 or R75130) gene
FT to form a GPI-anchored molecule, where the C-terminal peptide directs
FT attachment of a GPI anchor to the heterologous protein. The resulting
FT construct may be inserted in a vector for expression in a host cell.
FT The enzyme selectively hydrolyses the inositol-phosphate linkage of
FT GPI-anchored proteins, GPI lipids and related molecules, resulting in
FT direct secretion into the medium and simple recombinant protein
FT purification.
FT Sequence 839 AA:

```

```

Query Match 69.8%; Score 37; DB 16; Length 839;
Best Local Similarity 71.4%; Pred. No. 2.96e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 684 tlnqgs 690
OY 1 TLHQNGN 7

```

Search completed: Fri Apr 16 14:25:30 1999
Job time : 13 secs.

 W I D E R I N G (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPSEARCH protein - protein database search, using Smith-Waterman algorithm
 Run on: Fri Apr 16 14:22:33 1999; MPSEARCH time 2.66 Seconds
 Tabular output not generated. 171.661 Million cell updates/sec

Title: >US-09-011-797-3
 Description: (1-17) from US09011797.pep
 Perfect Score: 131
 Sequence: 1 FSEPMROYLVLSMOSQ 17

Scoring table:
 PAM 150
 Gap 15

Searched: 74019 seqs, 26840295 residues

Post-processing: Listing first 45 summaries

Database: swiss-prot35
 1:swissprot

Statistics: Mean 29.800; Variance 41.738; scale 0.714

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	131	100.0	176	1	PNOC_HUMAN NOCICEPTIN PRECURSOR (3.70e-15
2	131	100.0	181	1	PNOC_RAT NOCICEPTIN PRECURSOR (3.70e-15
3	131	100.0	187	1	PNOC_MOUSE NOCICEPTIN PRECURSOR (3.70e-15
4	66	50.4	468	1	CISY_GABEL PROBABLE CITRATE SYNTH	6.54e-01
5	65	49.6	317	1	ACCO_PEA 1-AMINOCYCLOPROPANE-1-	9.90e-01
6	64	48.9	207	1	YE33_METUA HYPOTHETICAL PROTEIN M	1.49e+00
7	62	47.3	320	1	ACC3_CUCME 1-AMINOCYCLOPROPANE-1-	3.35e+00
8	61	46.6	318	1	ACC1_CUCME 1-AMINOCYCLOPROPANE-1-	4.97e+00
9	61	46.6	733	1	MCMS_MOUSE DNA REPLICATION LICENSE	4.97e+00
10	61	46.6	734	1	MCMS_HUMAN DNA REPLICATION LICENSE	4.97e+00
11	61	46.6	894	1	YNS6_YEAST HYPOTHETICAL 99.5 KD P	4.97e+00
12	61	46.6	984	1	MYSA_DICDI MYOSIN 1A HEAVY CHAIN	4.97e+00
13	61	46.6	1189	1	PROTEIN-TYROSINE PHOSP	4.97e+00
14	60	45.8	144	1	YPNP_BACSV HYPOHETICAL 15.9 KD P	7.36e+00
15	60	45.8	422	1	CAN2_RABIT CALPAIN 2, LARGE (CATA	7.36e+00
16	60	45.8	700	1	CAN2_MOUSE CALPAIN 2, LARGE (CATA	7.36e+00
17	60	45.8	700	1	CAN2_RAT CALPAIN 2, LARGE (CATA	7.36e+00
18	60	45.8	780	1	CAN2_HUMAN CALPAIN-2, LARGE (CATA	7.36e+00
19	60	45.8	1187	1	PTNE_HUMAN PROTEIN-TYROSINE PHOSP	7.36e+00
20	60	45.8	4466	1	DYHC_ANTCR DYNEIN BETA CHAIN, CIL	7.36e+00
21	60	45.8	4466	1	DYHC_ANTCR DYNEIN BETA CHAIN, CIL	7.36e+00
22	59	45.0	89	1	RFAP_SALTY LIPOPOLYSACCHARIDE COR	1.08e+01
23	59	45.0	321	1	ACCO_DIACA PROBABLE 1-AMINOCYCLOP	1.08e+01

24	59	45.0	367	1	TRMU_MYCGE PROBABLE TRNA (5-METHY	1.08e+01
25	59	45.0	370	1	TRMU_MYCPN PROBABLE TRNA (5-METHY	1.08e+01
26	59	45.0	433	1	CISY_CHICK CITRATE SYNTHASE, MITO	1.08e+01
27	59	45.0	433	1	ENOG_HUMAN GAMMA ENOLASE (EC 4.2.	1.08e+01
28	59	45.0	464	1	CISY_PIG CITRATE SYNTHASE, MITO	1.08e+01
29	59	45.0	716	1	MCMS_XENLA DNA REPLICATION LICENSE	1.08e+01
30	59	45.0	946	1	GLNE_ECOLI GLUTAMATE-AMMONIA-LIGA	1.08e+01
31	58	44.3	315	1	ACC3_LYCES 1-AMINOCYCLOPROPANE-1-	1.59e+01
32	58	44.3	316	1	ACC2_LYCES 1-AMINOCYCLOPROPANE-1-	1.59e+01
33	58	44.3	319	1	ACCI_PETHY 1-AMINOCYCLOPROPANE-1-	1.59e+01
34	58	44.3	320	1	ACCG_PETHY 1-AMINOCYCLOPROPANE-1-	1.59e+01
35	58	44.3	371	1	IDH_METJA ISOCITRATE DEHYDROGENA	1.59e+01
36	58	44.3	393	1	FTSW_STINT3 PROBABLE CELL DIVISION	1.59e+01
37	58	44.3	433	1	ENOG_RAT GAMMA ENOLASE (EC 4.2.	1.59e+01
38	58	44.3	433	1	ENOG_MOUSE GAMMA ENOLASE (EC 4.2.	1.59e+01
39	58	44.3	486	1	NAM9_YEAST NAM9 PROTEIN PRECURSOR	1.59e+01
40	58	44.3	531	1	PEPE_ASPNG SERINE-TYPE CARBOXYPEP	1.59e+01
41	58	44.3	714	1	CAN1_HUMAN CALPAIN 1, LARGE (CATA	1.59e+01
42	58	44.3	1135	1	NINS_DROME NINAC SHORT PROTEIN (E	1.59e+01
43	58	44.3	1246	1	NARG_ECOLI RESPIRATORY NITRATE RE	1.59e+01
44	58	44.3	1501	1	NINL_DROME NINAC LONG PROTEIN (EC	1.59e+01
45	57	43.5	700	1	CAN2_CHICK CALPAIN 2, LARGE (CATA	2.31e+01

ALIGNMENTS

RESULT ID	1	PNOC_HUMAN	STANDARD:	PRT:	176 AA.
AC	013519;				
DT	01-NOV-1997 (REL. 35, LAST CREATED)				
DT	01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)				
DT	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)				
DE	NOCICEPTIN PRECURSOR (ORPHANIN FQ) (PNOC).				
GN	PNOC OR OFO.				
OS	HOMO SAPIENS (HUMAN).				
OC	EUTHERIA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;				
CC	EUTHERIA; PRIMATES.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE: 96323281.				
RA	MOLLEREAU C., SIMONS M.-J., SOULARUE P., LINERS F., VASSART G.,				
RA	MEUNIER J.-C., PARMENTIER M.,				
RL	PROC. NATL. ACAD. SCI. U.S.A. 93:8666-8670(1996).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE-BRAIN;				
RA	NOTACKER H.P., REINSCHIED R.R., MANSOUR A., HENNINGSEN R.A.,				
RA	MOMMA F.J. JR., WATSON S.J., CIVELLI O.,				
RL	SUBMITTED (FEB-1996) TO EMBL/GENBANK/DBD DATA BANKS.				
RN	[3]				
RP	STRUCTURE BY NMR OF NOCICEPTIN.				
RX	MEDLINE: 97312464.				
RA	SALVADORI S., PICONE D., TANGREDI T., GUERRINI R., SPADACCINI R.,				
RA	LAZARUS L.H., REGOLI D., TEMUSSI P.A.,				
RL	BIOCHEM. BIOPHYS. RES. COMMON. 233:640-643(1997).				
CC	-1- FUNCTION: NOCICEPTIN IS THE LIGAND OF THE OPIOID RECEPTOR-LIKE				
CC	RECEPTOR (OPRL1). IT MAY ACT AS A TRANSMITTER IN THE BRAIN BY				
CC	MODULATING NOCICEPTIVE AND LOCOMOTOR BEHAVIOR. MAY BE INVOLVED				
CC	IN NEURONAL DIFFERENTIATION AND DEVELOPMENT (BY SIMILARITY).				
CC	-1- SUBCELLULAR LOCATION: SECRETED INTO THE ENDOPLASMIC RETICULUM.				
CC	-1- TISSUE SPECIFICITY: PREDOMINANTLY IN THE BRAIN AND SPINAL CORD.				
CC	-1- PTM: SPECIFIC ENZYMAIC CLEAVAGES AT PAIRED BASIC RESIDUES				
CC	PROBABLY YIELD OTHER ACTIVE PEPTIDES BESIDES NOCICEPTIN.				
CC	-1- PTM: THE N-TERMINAL CONTAINS 6 CONSERVED CYSTEINES THOUGHT TO				
CC	BE INVOLVED IN DISULFIDE BONDING AND/OR PROCESSING.				
CC	-1- SIMILARITY: BELONGS TO THE OPIOIDS NEUROPEPTIDES PRECURSORS				
CC	FAMILY.				
DR	EMBL: X97370; E244777; -				
DR	EMBL: X97367; E244785; -				
DR	EMBL: X97368; E244785; JOINED.				
DR	EMBL: U48263; G1185010; -				
DR	MM: 601459; -				
DR	PROSITE: PS01252; OPIOIDS_PRECURSOR; 1.				

KW OPIOID PEPTIDE: NEUROTRANSMITTER: NEUROPEPTIDE;
 KM CLEAVAGE ON PAIR OF BASIC RESIDUES: SIGNAL.
 FT SIGNAL 1 19 POTENTIAL.
 FT PEPTIDE 98 127 NEUROPEPTIDE 1 (PROBABLE).
 FT PEPTIDE 130 146 NOCICEPTIN (ORPHANIN FQ).
 FT PEPTIDE 149 165 NEUROPEPTIDE 2 (PROBABLE).
 SQ SEQUENCE 176 AA: 20295 MW: 8055843A CRC32;

Query Match 100.0%; Score 131; DB 1; Length 176;
 Best Local Similarity 100.0%; Pred. No. 3,70e-15;
 Matches 17: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 149 FSEFMROYLVLSMOSSQ 165
 1 FSEFMROYLVLSMOSSQ 17

RESULT 2
 ID PNCOC RAT STANDARD: PRT: 181 AA.
 AC 062923: 064162: (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DE NOCICEPTIN PRECURSOR (ORPHANIN FQ) (PNCOC) (ORL1 RECEPTOR AGONIST).
 GN PNCOC.
 OS RATTUS NORVEGICUS (RAT).
 OC EUTHERIA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 RN EUTHERIA; RODENTIA.
 RC SEQUENCE FROM N.A.
 RA STRAIN-SPRAGUE-DAWLEY;
 RA NOTACKER H.P., REINSCHIED R.K., MANSOUR A., HENNINGSEN R.A.,
 RA MONSMA F.J. JR., MATSON S.J., CIVELLI O.,
 RL SUBMITTED (FEB-1996) TO EMBL/GENBANK/DBD DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 96323281.
 RA MOLLEREAU C., SIMONS M.-J., SOULARUE P., LINERS F., VASSART G.,
 RA MEUNIER J.-C., PARMENTIER M.,
 RL PROC. NATL. ACAD. SCI. U.S.A. 93:8666-8670(1996).
 RN [3]
 RP SEQUENCE OF 13-181 FROM N.A., AND SEQUENCE OF 135-151.
 RX TISSUE-BRAIN;
 RA MEDLINE: 96013745.
 RA MEUNIER J.-C., MOLLEREAU C., TOLL L., STAUDEAU C., MOISAND C.,
 RA ALVIERIE P., BUROU J.-L., GUILLOT J.-C., FERRARA P.,
 RA MONSARAT B., MARAGUIL H., VASSART G., PARMENTIER M., COSTENTIN J.,
 RL NATURE 377:532-535(1995).
 CC -1- FUNCTION: NOCICEPTIN IS THE LIGAND OF THE OPIOID RECEPTOR-LIKE
 RECEPTOR (OPRL1). IT MAY ACT AS A TRANSMITTER IN THE BRAIN BY
 MODULATING NOCICEPTIVE AND LOCOMOTOR BEHAVIOR. MAY BE INVOLVED
 IN NEURONAL DIFFERENTIATION AND DEVELOPMENT.
 CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN THE SPINAL CORD
 AND BRAIN, BEING MORE ABUNDANT IN THE HYPOTHALAMUS AND STRIATUM.
 CC -1- ALSO FOUND IN SMALL AMOUNTS IN OVARY.
 CC -1- PTH: SPECIFIC ENZYMAIC CLEAVAGES AT PAIRED BASIC RESIDUES
 PROBABLY YIELD OTHER ACTIVE PEPTIDES BESIDES NOCICEPTIN.
 CC -1- PTH: THE N-TERMINAL CONTAINS 6 CONSERVED CYSTEINES THOUGHT TO
 BE INVOLVED IN DISULFIDE BONDING AND/OR PROCESSING.
 CC -1- SIMILARITY: BELONGS TO THE OPIOIDS NEUROPEPTIDES PRECURSORS
 FAMILY.
 CC EMBL: U48262; G1185012; -.
 DR EMBL: X97375; E244795; -.
 DR EMBL: S79730; G117239; -.
 DR PROSITE: PS01252; OPIOIDS.PRECURSOR; 1.
 KM OPIOID PEPTIDE: NEUROTRANSMITTER: NEUROPEPTIDE;
 KM CLEAVAGE ON PAIR OF BASIC RESIDUES: SIGNAL; REPEAT.
 FT SIGNAL 1 19 POTENTIAL.
 FT PEPTIDE 98 132 NEUROPEPTIDE 1 (PROBABLE).
 FT PEPTIDE 135 151 NOCICEPTIN (ORPHANIN FQ).
 FT PEPTIDE 154 170 NEUROPEPTIDE 2 (PROBABLE).
 FT DOMAIN 109 120 2 x 6 AA TANDEM REPEATS OF D-A-E-P-X-A.

FT REPEAT 109 114 1.
 FT REPEAT 115 120 2.
 SQ SEQUENCE 181 AA: 20478 MW: 1640366F CRC32;

Query Match 100.0%; Score 131; DB 1; Length 181;
 Best Local Similarity 100.0%; Pred. No. 3,70e-15;
 Matches 17: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 154 FSEFMROYLVLSMOSSQ 170
 1 FSEFMROYLVLSMOSSQ 17

RESULT 3
 ID PNCOC MOUSE STANDARD: PRT: 187 AA.
 AC 064387: 061105: 061938:
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DE NOCICEPTIN PRECURSOR (ORPHANIN FQ) (PNCOC) (N23K / N27K).
 GN PNCOC OR NPNC1.
 OS MUS MUSCULUS (MOUSE).
 OC EUMAROTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 RN EUTHERIA; RODENTIA.
 RC SEQUENCE FROM N.A.
 RX TISSUE-BRAIN;
 RA MEDLINE: 96216718.
 RA HOOTANT T., NISHI M., TAKESHIMA H., NUKADA T., SUGIMOTO T.,
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 219:714-719(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 96106851.
 RA SAITO Y., MARUYAMA K., SAIDO T.C., KAWASHIMA S.,
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 217:539-545(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-129/SV;
 RX MEDLINE: 96323281.
 RA MOLLEREAU C., SIMONS M.-J., SOULARUE P., LINERS F., VASSART G.,
 RA MEUNIER J.-C., PARMENTIER M.,
 RL PROC. NATL. ACAD. SCI. U.S.A. 93:8666-8670(1996).
 RN [4]
 RP SEQUENCE FROM N.A. (LONG FORM).
 RX MEDLINE: 96279082.
 RA SAITO Y., MARUYAMA K., KAWANO H., HAGINO-YAMAGISHI K.,
 RA KAWAMURA K., SAIDO T.C., KAWASHIMA S.,
 RL J. BIOL. CHEM. 271:15615-15622(1996).
 RN [5]
 RP SEQUENCE OF 79-187 FROM N.A.
 RX TISSUE-BRAIN;
 RA MEDLINE: 96207555.
 RA PAN Y.-X., XU J., PASTERNAK G.W.,
 RL BIOCHEM. J. 315:11-13(1996).
 CC -1- FUNCTION: NOCICEPTIN IS THE LIGAND OF THE OPIOID RECEPTOR-LIKE
 RECEPTOR (OPRL1). IT MAY ACT AS A TRANSMITTER IN THE BRAIN BY
 MODULATING NOCICEPTIVE AND LOCOMOTOR BEHAVIOR. MAY BE INVOLVED
 IN NEURONAL DIFFERENTIATION AND DEVELOPMENT. WHEN ADMINISTERED
 INTRACEREBROVENTRICULARLY, NOCICEPTIN INDUCES HYPERALGESIA AND
 DECREASES LOCOMOTOR ACTIVITY.
 CC -1- SUBCELLULAR LOCATION: SECRETED INTO THE ENDOPLASMIC RETICULUM.
 CC -1- TISSUE SPECIFICITY: BRAIN AND SPINAL CORD. LOW LEVELS IN
 KIDNEY AND SPLEEN.
 CC -1- DEVELOPMENTAL STAGE: IN EMBRYONIC BRAIN, FIRST DETECTED AT DAY 14
 AND IN POSTNATAL BRAIN, LEVELS INCREASE IN DAY 1 AND DAY 18.
 CC -1- LEVELS DECREASE SIGNIFICANTLY IN ADULTS.
 CC -1- PTH: SPECIFIC ENZYMAIC CLEAVAGES AT PAIRED BASIC RESIDUES
 PROBABLY YIELD OTHER ACTIVE PEPTIDES BESIDES NOCICEPTIN.
 CC -1- PTH: THE N-TERMINAL CONTAINS 6 CONSERVED CYSTEINES THOUGHT TO
 BE INVOLVED IN DISULFIDE BONDING AND/OR PROCESSING.
 CC -1- ALTERNATIVE PRODUCTS: A LONGER FORM, N27K IS PROBABLY PRODUCED
 BY ALTERNATIVE SPLICING.
 CC -1- SIMILARITY: BELONGS TO THE OPIOIDS NEUROPEPTIDES PRECURSORS

KM HYPOTHETICAL PROTEIN, HYDROLASE, NUCLEASE, ENDONUCLEASE.
 FT ACT_SITE 70 70 BY SIMILARITY.
 FT ACT_SITE 78 78 BY SIMILARITY.
 FT ACT_SITE 148 148 BY SIMILARITY.
 SQ SEQUENCE 207 AA; 24547 MW; E0CF3191 CRC32;
 Query Match 48.9%; Score 64; DB 1; Length 207;
 Best Local Similarity 60.0%; Pred. No. 1.49e+00;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 Db 6 IMRKLISM 15
 Oy 4 FMROYLVLSM 13
 RESULT 7
 ID ACC3_CUCME STANDARD; PRT; 320 AA.
 AC P54847;
 DT 01-OCT-1996 (REL. 34, CREATED)
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
 DE 1-AMINOCYCLOPROPANE-1-CARBOXYLATE OXIDASE 3 (ACC OXIDASE) (ETHYLENE-FORMING ENZYME) (EFE).
 DE CUCUMIS MELO (MUSKMELO).
 GN AC03.
 OS CUCUMIS MELO (MUSKMELO).
 OC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE;
 CC VIOLAE; CUCURBITACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. CANTALOUPE CHARENTAIS; TISSUE-LEAF;
 RX MEDLINE; 96213994.
 RA LASSERE E., BOUQUIN T., HERNANDEZ J.A., BULL J., PECH J.C.,
 BALAGUE C.;
 RL MOL. GEN. GENET. 251:81-90(1996).
 CC -1- CATALYTIC ACTIVITY: 1-AMINOCYCLOPROPANE-1-CARBOXYLATE + O2 =
 CC ETHYLENE + HCN + CO(2) + 2 H(2)O.
 CC -1- COFACTOR: IRON AND ASCORBATE.
 CC -1- PATHWAY: LAST STEP IN THE BIOSYNTHESIS OF ETHYLENE.
 CC -1- TISSUE SPECIFICITY: FLOWERS.
 CC -1- SIMILARITY: BELONGS TO THE IRON/ASCORBATE-DEPENDENT FAMILY OF
 CC OXIDOREDUCTASES.
 DR EMBL; X95553; E221407; -
 KW FRUIT RIPENING; ETHYLENE BIOSYNTHESIS; OXIDOREDUCTASE; IRON;
 KM VITAMIN C; MULTIGENE FAMILY.
 SQ SEQUENCE 320 AA; 36397 MW; F3FD4947 CRC32;
 Query Match 47.3%; Score 62; DB 1; Length 320;
 Best Local Similarity 29.4%; Pred. No. 3.35e+00;
 Matches 5; Conservative 6; Mismatches 6; Indels 0; Gaps 0;
 Db 282 FDDYMKLVGLFQAKE 298
 Oy 1 FSEFMROYLVLSMOSSQ 17
 RESULT 8
 ID ACC1_CUCME STANDARD; PRT; 318 AA.
 AC Q04644;
 DT 01-OCT-1993 (REL. 27, CREATED)
 DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
 DE 1-AMINOCYCLOPROPANE-1-CARBOXYLATE OXIDASE 1 (ACC OXIDASE) (ETHYLENE-FORMING ENZYME) (EFE) (PME1).
 DE CUCUMIS MELO (MUSKMELO).
 GN AC01.
 OS CUCUMIS MELO (MUSKMELO).
 OC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE;
 CC VIOLAE; CUCURBITACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. CANTALOUPE CHARENTAIS; TISSUE-FRUIT;
 RX MEDLINE; 93185633.
 RA BALAGUE C., WATSON C.F., TURNER A.J., ROUGE P., PICTON S.,
 PECH J.C., GRIERSON D.;

RL EUR. J. BIOCHEM. 212:27-34(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. ANDES; TISSUE-FRUIT;
 RA YAMAMOTO M., MIKI T., ISHII Y., NAKAGAWA H., OGURA N., SATO T.;
 RN SUBMITTED (MAR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. CANTALOUPE CHARENTAIS; TISSUE-LEAF;
 RX MEDLINE; 96213994.
 RA LASSERE E., BOUQUIN T., HERNANDEZ J.A., BULL J., PECH J.C.,
 BALAGUE C.;
 RL MOL. GEN. GENET. 251:81-90(1996).
 CC -1- CATALYTIC ACTIVITY: 1-AMINOCYCLOPROPANE-1-CARBOXYLATE + O2 =
 CC ETHYLENE + HCN + CO(2) + 2 H(2)O.
 CC -1- COFACTOR: IRON AND ASCORBATE.
 CC -1- PATHWAY: LAST STEP IN THE BIOSYNTHESIS OF ETHYLENE.
 CC -1- TISSUE SPECIFICITY: FRUIT.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING FRUIT RIPENING.
 CC -1- SIMILARITY: BELONGS TO THE IRON/ASCORBATE-DEPENDENT FAMILY OF
 CC OXIDOREDUCTASES.
 DR EMBL; X69935; G22663; -
 DR EMBL; D31727; G695400; -
 DR EMBL; X95551; E221404; -
 DR PIR; S29395; S29395.
 KW FRUIT RIPENING; ETHYLENE BIOSYNTHESIS; OXIDOREDUCTASE; IRON;
 KM VITAMIN C; MULTIGENE FAMILY.
 SQ SEQUENCE 318 AA; 36127 MW; 469A2496 CRC32;
 Query Match 46.6%; Score 61; DB 1; Length 318;
 Best Local Similarity 29.4%; Pred. No. 4.97e+00;
 Matches 5; Conservative 6; Mismatches 6; Indels 0; Gaps 0;
 Db 284 FEDYMKLVGLFQAKE 300
 Oy 1 FSEFMROYLVLSMOSSQ 17
 RESULT 9
 ID MCM5_MOUSE STANDARD; PRT; 733 AA.
 AC P49718;
 DT 01-OCT-1996 (REL. 34, CREATED)
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE DNA REPLICATION LICENSING FACTOR MCM5 (CDC46 HOMOLOG) (P1-CDC46).
 GN MCM5 OR MCM5 OR CDC46.
 OS MUS MUSCULUS (MOUSE).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 CC EUTHERIA; RODENTIA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE; 95334361.
 RA KIMURA H., TAKIZAWA N., NOZAKI N., SUGIMOTO K.;
 RL NUCLEIC ACIDS RES. 23:2097-2104(1995).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- SIMILARITY: BELONGS TO THE MCM FAMILY.
 DR EMBL; D26090; G940404; -
 DR MGD; MGI:103197; MCM5.
 DR PROSITE; PS00847; MCM.1; 1.
 DR PROSITE; PS50051; MCM.2; 1.
 KW TRANSCRIPTION REGULATION; DNA-BINDING; NUCLEAR PROTEIN;
 KM DNA REPLICATION; ATP-BINDING.
 FT DOMAIN 330 536 MCM.
 FT NP_BIND 380 387 ATP (POTENTIAL).
 SQ SEQUENCE 733 AA; 82342 MW; DA793E3F CRC32;
 Query Match 46.6%; Score 61; DB 1; Length 733;
 Best Local Similarity 70.0%; Pred. No. 4.97e+00;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 Db 35 FKEFLROYRY 44
 Oy 1 FSEFMROYLV 10

RESULT 10
ID MCM5_HUMAN STANDARD; PRT; 734 AA.
AC P33992; Q14578;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DE DNA REPLICATION LICENSING FACTOR MCM5 (CDC46 HOMOLOG) (P1-CDC46).
GN MCM5 OR CDC46.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RA HU B.;
RL SUBMITTED (JUL-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RM [2]
RP SEQUENCE FROM N.A.
RA MINUOTA S., NISHIMOTO S., KUBOTA Y., TAKISAWA H., NOJIMA H.;
RL SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE OF 352-590 FROM N.A.
RC TISSUE-CERVIX;
RX MEDLINE: 94089373.
RA HU B., BURKHART R., SCHULTE D., MUSANL C., KNIPPERS R.;
RL NUCLEIC ACIDS RES. 21:5289-5293(1993).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: BELONGS TO THE MCM FAMILY.
DR EMBL: X74795; G895843; -
DR EMBL: D83986; G1237079; -
DR PIR: S43199; S43199.
DR PROSITE: PS00847; MCM_1; 1.
DR PROSITE: PS50051; MCM_2; 1.
KM TRANSCRIPTION REGULATION: DNA-BINDING; NUCLEAR PROTEIN;
KM DNA REPLICATION: ATP-BINDING.
FT DOMAIN 331 537 MCM.
FT NP_BIND 381 388 ATP (POTENTIAL).
FT CONFLICT 180 180 S -> T (IN REF. 2).
FT CONFLICT 241 241 V -> M (IN REF. 2).
FT CONFLICT 434 434 W -> M (IN REF. 2).
FT CONFLICT 527 527 E -> V (IN REF. 2).
FT CONFLICT 593 603 AROHERSDRR -> PSTSTGYTA (IN REF. 2).
FT CONFLICT 612 613 QL -> HV (IN REF. 1).
FT CONFLICT 717 717 R -> P (IN REF. 1).
SQ SEQUENCE 734 AA; 82308 MM; 3D6479B2 CRC32;
Query Match 46.6%; Score 61; DB 1; Length 734;
Best Local Similarity 70.0%; Pred. No. 4.97e+00;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

DB 35 EKEFLQRYV 44
Y 1 1 1 1 1
OY 1 FSEFMROYLV 10

RESULT 11
ID YN86_YEAST STANDARD; PRT; 894 AA.
AC P27514;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE HYPOHETICAL 99.5 KD PROTEIN IN URK1-SM1 INTERGENIC REGION.
GN YNR013C OR N2052.
OS SACCAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / FY1679;
RX MEDLINE: 95208356.
RA VERHAASSELT P., AERT R., VOET M., VOLCKAERT G.;
RL YEAST 10:1355-1361(1994).
RN [2]

RP SEQUENCE OF 1-293 FROM N.A.
RA MAURER C.T.C., URBANUS J.H.M., PLANTA R.J.;
RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE OF 784-894 FROM N.A.
RC STRAIN-ATCC 28383 / FL100;
RX MEDLINE: 90384830.
RA KERN L.;
RL NUCLEIC ACIDS RES. 18:5279-5279(1990).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE NAD(P)H/O87 FAMILY OF TRANSPORTERS.
CC PHO87 SUBFAMILY.
DR EMBL: X77395; G496729; -
DR EMBL: Z71628; E239581; -
DR EMBL: Z71629; E328814; -
DR EMBL: X53998; E32709; ALT_INIT.
DR PIR: S29375; S29375.
DR PIR: S45135; S45135.
KM HYPOHETICAL PROTEIN: TRANSMEMBRANE.
FT TRANSMEM 430 450 POTENTIAL.
FT TRANSMEM 474 494 POTENTIAL.
FT TRANSMEM 511 531 POTENTIAL.
FT TRANSMEM 557 577 POTENTIAL.
FT TRANSMEM 602 622 POTENTIAL.
FT TRANSMEM 642 662 POTENTIAL.
FT TRANSMEM 682 702 POTENTIAL.
FT TRANSMEM 706 726 POTENTIAL.
FT TRANSMEM 738 758 POTENTIAL.
FT TRANSMEM 777 797 POTENTIAL.
FT TRANSMEM 799 819 POTENTIAL.
FT TRANSMEM 824 844 POTENTIAL.
FT TRANSMEM 874 894 POTENTIAL.
SQ SEQUENCE 894 AA; 99490 MM; 949FB214 CRC32;
Query Match 46.6%; Score 61; DB 1; Length 894;
Best Local Similarity 46.2%; Pred. No. 4.97e+00;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

DB 735 FNNFMWTVLVAM 747
Y 1 1 1 1 1
OY 1 FSEFMROYLVLM 13

RESULT 12
ID MYSA_DICDI STANDARD; PRT; 994 AA.
AC P22467;
DT 01-AUG-1991 (REL. 19, CREATED)
DT 01-NOV-1991 (REL. 19, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE MYOSIN IA HEAVY CHAIN (MYOSIN-LIKE PROTEIN ABMA).
GN ABMA OR DMIA.
OS DICTYOSTELIUM DISCOIDEUM (SLIME MOLD).
OC EUKARYOTA; PROTOZOA; SARCOMASTIGOPHORA; SARCODINA; RHIZOPODA;
OC EUMYCETOTOA; DICTYOSTELIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A5-2;
RX MEDLINE: 92096486.
RA TITUS M.A., WARRICK H.M., SPUDICH J.A.;
RL CELL REGUL. 1:55-63(1989).
CC -1- FUNCTION: ACTIN-BASED MOTOR PROTEIN, POSSIBLY INVOLVED IN A WIDE RANGE OF MOTILE PROCESSES, SUCH AS CELL MOVEMENT ACROSS A SURFACE, AND EXTENSION AND RETRACTION OF PSEUDOPODIA OR LAMELLIPODIA.
CC -1- SUBUNIT: MYOSIN I HEAVY CHAIN IS SINGLE-HEADED. DIMER OF A HEAVY AND A LIGHT CHAIN. INABILITY TO SELF-ASSEMBLE INTO FILAMENTS.
CC -1- SIMILARITY: BELONGS TO THE MYOSIN TYPE I FAMILY (SMALL MYOSINS).
DR EMBL: S73909; G241268; -
DR PIR: J00151; J00151.
DR DICTYDB: DDO1028; ABMA.
KM MYOSIN: ACTIN-BINDING; ATP-BINDING; MULTIGENE FAMILY.
FT DOMAIN 1 782 GLOBULAR HEAD-LIKE DOMAIN.
FT NP_BIND 105 112 NON ALPHA-HELICAL, C-TERMINAL DOMAIN.
ATP (POTENTIAL).

FT	DOMAIN	575	655	ACTIN-BINDING.
FT	DOMAIN	830	931	MEMBRANE-BINDING (POTENTIAL).
SO	SEQUENCE	994 AA:	113286 MW;	5609233f CRC32:
Query Match				
Best Local Similarity		46.6%;	Score 61;	DB 1; Length 994;
Matches		6; Conservative	5;	Mismatches 2; Indels 0; Gaps 0;
Db	653 FSEKMRYYMICK	665		
QY	1 FSEFMRYLVLSM	13		
RESULT 13				
ID	PTNE.MOUSE	STANDARD:	PRT:	1189 AA.
AC	062130:			
DT	01-NOV-1997 (REL. 35, CREATED)			
DT	01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)			
DT	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)			
DE	PROTEIN-TYROSINE PHOSPHATASE PTP36 (EC 3.1.3.48).			
GN	PTPN14.			
OS	MUS MUSCULUS (MOUSE).			
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;			
OC	EUTHERIA; RODENTIA.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CB-17-SCID; TISSUE=THYMUS;			
RX	MEDLINE; 94354845.			
RA	SANADA M., OGATA M., FUJINO Y., HANAOKA T.;			
RL	BIOCHEM. BIOPHYS. RES. COMMUN. 203:479-484(1994).			
CC	- FUNCTION: MAY BE INVOLVED IN THE REGULATION OF T CELL DEVELOPMENT.			
CC	- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =			
CC	- TISSUE SPECIFICITY: THYMUS; IN CELLS OF BOTH HEMATOPOIETIC AND			
CC	- NON-HEMATOPOIETIC ORIGINS.			
CC	- SIMILARITY: CONTAINS A DOMAIN FOUND IN BAND 4.1, EKRIN, MOESIN,			
CC	- RADIAXN. AND TALIN.			
CC	- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-			
CC	- TYROSINE PHOSPHATASE FAMILY.			
DR	EMBL; D31842; G507331; -			
DR	MGI; D31842; G507331; -			
DR	PROSITE; PS00660; BAND_41.1; 1.			
DR	PROSITE; PS00661; BAND_41.2; 1.			
DR	PROSITE; PS00657; BAND_41.3; 1.			
DR	PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.			
DR	PROSITE; PS50055; TYR_PHOSPHATASE_PP; 1.			
DR	PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.			
KW	STRUCTURAL PROTEIN; CYTOSKELETON; HYDROLASE.			
FT	DOMAIN	75	239	BAND 4.1-LIKE DOMAIN.
FT	DOMAIN	935	1189	PROTEIN-TYROSINE PHOSPHATASE.
FT	ACT_SITE	1123	1173	BY SIMILARITY.
FT	DOMAIN	566	573	POLY-PRO.
FT	DOMAIN	635	639	POLY-GLY.
FT	DOMAIN	712	718	POLY-GLU.
SO	SEQUENCE	1189 AA;	135030 MW;	CDD0FEC4 CRC32;
Query Match				
Best Local Similarity		46.6%;	Score 61;	DB 1; Length 1189;
Matches		5; Conservative	7;	Mismatches 0; Indels 0; Gaps 0;
Db	160 EFLREYVLEPPMD	171		
QY	3 EFMRYLVLSMQ	14		
RESULT 14				
ID	YHPH_BACSU	STANDARD:	PRT:	144 AA.
AC	P54170;			
DT	01-OCT-1996 (REL. 34, CREATED)			
DT	01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)			
DT	01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)			
DE	HYPOTHETICAL 15.9 KD PROTEIN IN ILVD-THYB INTERGENIC REGION.			
GN	YHPH.			

OS	BACILLUS SUBTILIS.		
OC	PROCAROTIA: FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-168 / MARBURG;		
RA	SOROKIN A.V., AZEVEDO V., ZUMSTEIN E., GALLERON N., EHRLICH S.D.,		
RA	SERFOR P.;		
RL	SUBMITTED (APR-1996) TO EMBL/GENBANK/DDJ DATA BANKS.		
CC	-1- SIMILARITY: STRONG. TO B.SUBTILIS YQIW.		
DR	EMBL; L77246; G1256636; -		
DR	EMBL; Z99115; E1183633; -		
DR	SUBTILIST; BG11614; YPHF.		
KW	HYPOTHETICAL PROTEIN.		
SC	SEQUENCE 144 AA; 15875 MW; 18864F3F CRC32;		
<hr/>			
Query Match	45.8%;	Score 60;	DB 1; Length 144;
Best Local Similarity	50.0%;	Pred. No. 7.36e+00;	
Matches	7; Conservative	5; Mismatches	1; Indels 1; Gaps
<hr/>			
Dd	5 YEERYMQ-LVYPMR 17 : : : : 1 FSEFMRYGLVLSMQ 14		
<hr/>			
RESULT 15	CAN2_RABIT STANDARD; FTI; 422 AA.		
ID	P06814;		
AC	01-JAN-1988 (REL. 06, CREATED)		
DT	01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)		
DT	01-FEB-1986 (REL. 33, LAST ANNOTATION UPDATE)		
MT	CALPAIN 2, LARGE (CATALYTIC) SUBUNIT (EC 3.4.22.17) (CALCIUM-ACTIVATED		
DE	NEUTRAL PROTEINASE) (CANP) (M-TYPE) (FRAGMENT).		
GN	CANP2.		
OS	ORYZOTOLAGUS CUNICULUS (RABBIT).		
OC	EUKARYOTA; METAOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;		
OC	EUTHERIA; LANGOMORPHA.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	EMBL; 86250902.		
RA	EMORI Y., KAWASAKI H., SUGIHARA H., IMAJOH S., KAWASHIMA S.,		
RA	SUZUKI K.;		
RL	J. BIOL. CHEM. 261:9465-9471(1986).		
RN	[2]		
RP	CALCIUM-BINDING DATA.		
RC	MEDLINE; 87279982.		
RA	MINAMI Y., EMORI Y., KAWASAKI H., SUZUKI K.;		
RL	J. BIOCHEM. 101:889-895(1987).		
CC	-1- FUNCTION: CALPAINS ARE CALCIUM-ACTIVATED NON-LYSOSOMAL THIO-		
CC	PROTEASES.		
CC	-1- ENZYME REGULATION: CALPAIN II IS ACTIVATED BY MITOLAR		
CC	CONCENTRATIONS OF CALCIUM.		
CC	-1- SUBUNIT: HETERODIMER OF A LARGE (CATALYTIC) AND A SMALL		
CC	(REGULATOR) SUBUNIT.		
CC	-1- SUBCELLULAR LOCATION: CYTOPLASMIC.		
CC	-1- IN MAMMALS THERE SEEM TO BE 3 DISTINCT FORMS: CALPAIN I (MICHO-		
CC	MOLES CA++ REQUIRING), CALPAIN II (MULTI-MOLE CA++ REQUIRING),		
CC	AND CALPAIN P94. THE SMALL UNIT IS COMMON TO ALL FORMS.		
CC	-1- THIS PROTEIN SEEMS TO BIND TWO MOLES OF CALCIUM.		
CC	-1- SIMILARITY: TO OTHER EF-HAND CALCIUM BINDING PROTEINS.		
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2; ALSO KNOWN AS THE		
CC	CALPAIN FAMILY OF THIOL PROTEASES.		
DR	EMBL; M13797; G165666; -		
DR	PIR; B24815; B24815		
DR	PROSITE; PS00018; EF_HAND; 2.		
DR	PROSITE; PS00139; THIOL_PROTEASE_CYS; PARTIAL.		
DR	PROSITE; PS00639; THIOL_PROTEASE_HIS; PARTIAL.		
DR	PROSITE; PS00640; THIOL_PROTEASE_ASN; PARTIAL.		
KW	HYDROLASE; THIOL PROTEASE; CALCIUM-BINDING; MULTIGENE FAMILY.		
FT	NTER		
FT	DOMAIN <1 49 1		
FT	DOMAIN 50 255 II, THIOL PROTEASE.		
FT	DOMAIN 256 422 III.		
FT	DOMAIN IV, CALCIUM-BINDING.		
FT	CA_BIND 307 318 SITE 1.		

FT	CA BIND	337	348	SITE 2.
FT	DOMAIN	372	383	ANCESTRAL CALCIUM SITE 3 (POTENTIAL).
FT	DOMAIN	402	413	ANCESTRAL CALCIUM SITE 4 (POTENTIAL).
SO	SEQUENCE	422 AA;	49494 MM;	BIFDFFA CRC32;

Query Match 45.88; Score 60; DB 1; Length 422;
Best Local Similarity 62.58; Pred. No. 7.36e+00;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db	50 FSDFLRHY	57
OY	1 FSEFMROY	8

Search completed: Fri Apr 16 14:22:41 1999
Job time : 8 secs.

THIS PAGE BLANK (USPTO)

 M P E S R E L I
 (TM)

Release 3.1A John F. Collins, Bioinformatics Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPearch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Fri Apr 16 14:26:23 1999: Maper time 2.36 Seconds
 Tabular output not generated. 90.843 Million cell updates/sec

Title: >US-09-011-797-4
 Description: (1-8) from US09011797.pep
 Perfect Score: 53
 Sequence: 1 TLHONGNV 8

Scoring table: PAM 150
 Gap 15

Searched: 74019 seqs, 26840295 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: swiss-prot35
 1:swissprot

Statistics: Mean 20.317; Variance 18.409; scale 1.104

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	53	100.0	176	1	PNOC_HUMAN NOCICEPTIN PRECURSOR (1.74e+03
2	53	100.0	181	1	PNOC_HUMAN NOCICEPTIN PRECURSOR (1.74e+03
3	53	100.0	187	1	PNOC_MOUSE NOCICEPTIN PRECURSOR (1.74e+03
4	45	84.9	1051	1	UBA2_MOUSE UBIQUITIN-ACTIVATING E	3.94e+01
5	45	84.9	1051	1	UBA1_WHEAT UBIQUITIN-ACTIVATING E	3.94e+01
6	44	83.0	384	1	TEAP_MOUSE HYPOTHETICAL 43.6 KD P	7.46e+01
7	43	81.1	377	1	SKN1_CANAL BETA-GLUCAN SYNTHESIS-	4.75e+00
8	41	77.4	224	1	SEPI_BACSU SFP PROTEIN.	4.75e+00
9	41	77.4	261	1	YIBA_ECOLI HYPOTHETICAL 28.7 KD P	4.75e+00
10	41	77.4	508	1	FMS1_YEAST FMS1 PROTEIN.	4.75e+00
11	41	77.4	726	1	PROBABLE SERINE/THREON	4.75e+00
12	41	77.4	863	1	STP1_YEAST STP1 PROTEIN.	4.75e+00
13	40	75.5	298	1	YDHE_ECOLI HYPOTHETICAL OXIDOREDU	8.61e+00
14	40	75.5	639	1	DNA GYRASE SUBUNIT B (8.61e+00
15	39	73.6	147	1	SP22_BACME ANTI-SIGMA F FACTOR (S	1.54e+01
16	39	73.6	232	1	PRAY_ECOLI LIPOPOLYSACCHARIDE COR	1.54e+01
17	39	73.6	232	1	PRAY_SALTY LIPOPOLYSACCHARIDE COR	1.54e+01
18	39	73.6	412	1	TADC_ECOLI HYPOTHETICAL FIMBRIN-L	1.54e+01
19	39	73.6	462	1	ANFK_AZOVI NITROGENASE IRON-IRON	1.54e+01
20	39	73.6	720	1	KRE6_YEAST BETA-GLUCAN SYNTHESIS-	1.54e+01
21	39	73.6	723	1	MY5B_MOUSE MYOSIN 5B (MYOSIN-LIKE	1.54e+01
22	39	73.6	740	1	KRE6_CANAL BETA-GLUCAN SYNTHESIS-	1.54e+01
23	39	73.6	749	1	CATA_ASPPU CATALASE A (EC 1.1.1.1.	1.54e+01

RESULT	ID	PNOC_HUMAN	STANDARD	PRT	176 AA.	ALIGNMENTS
AC	013519					
DT	01-NOV-1997	(REL. 35, CREATED)				
DT	01-NOV-1997	(REL. 35, LAST SEQUENCE UPDATE)				
DE	01-NOV-1997	(REL. 35, LAST ANNOTATION UPDATE)				
DE	NOCICEPTIN PRECURSOR (ORPHANIN FQ) (PNOC).					
GN	PNOC OR OFO.					
OS	HOMO SAPIENS (HUMAN).					
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;					
OC	EUETHERIA; PRIMATES.					
RN	(1)					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE: 96323281.					
RA	MOLLEREAU C., SIMONS M.-J., SOULARUE P., LINERS F., VASSART G.,					
RL	MEUNIER J.-C., PARENTIER M.,					
RL	PROC. NATL. ACAD. SCI. U.S.A. 93:8666-8670(1996).					
RN	(2)					
RP	SEQUENCE FROM N.A.					
RC	TISSUE-BRAIN;					
RA	NOTHACKER H.P., REINSCHIED R.K., MANSOUR A., HENNINGSSEN R.A.,					
RL	MONSMA F.J.J., JR., WATSON S.J., CIVELLI O.,					
RL	SUBMITTED (FEB-1996) TO EMBL/GENBANK/DBJ DATA BANKS.					
RN	(3)					
RP	STRUCTURE BY NMR OF NOCICEPTIN.					
RX	MEDLINE: 97312464.					
RA	SALVADORI S., PICONI D., TANCREDI T., GUERRINI R., SPADACCINI R.,					
RL	LAZARUS L.H., REGOLI D., TEMDSSI P.A.,					
RL	BIOCHEM. BIOPHYS. RES. COMMON. 233:640-643(1997).					
CC	- FUNCTION: NOCICEPTIN IS THE LIGAND OF THE OPIOID RECEPTOR-LIKE					
CC	RECEPTOR (OPR1). IT MAY ACT AS A TRANSMITTER IN THE BRAIN BY					
CC	MODULATING NOCICEPTIVE AND LOCOMOTOR BEHAVIOR. MAY BE INVOLVED					
CC	IN NEURONAL DIFFERENTIATION AND DEVELOPMENT (BY SIMILARITY).					
CC	- SUBCELLULAR LOCATION: SECRETED INTO THE ENDOPLASMIC RETICULON.					
CC	- TISSUE SPECIFICITY: PREDOMINANTLY IN THE BRAIN AND SPINAL CORD.					
CC	- PTM: SPECIFIC ENZYMOLOGICAL CLEAVAGES AT PAIRED BASIC RESIDUES					
CC	PROBABLY YIELD OTHER ACTIVE PEPTIDES BESIDES NOCICEPTIN.					
CC	- PTM: THE N-TERMINAL CONTAINS 6 CONSERVED CYSTEINES THOUGHT TO					
CC	BE INVOLVED IN DISULFIDE BONDING AND/OR PROCESSING.					
CC	- SIMILARITY: BELONGS TO THE OPIOIDS NEUROPEPTIDES PRECURSORS					
CC	FAMILY.					
DR	EMBL: X97367; E244777; -					
DR	EMBL: X97367; E244785; -					
DR	EMBL: X97368; E244785; JOINED.					
DR	EMBL: U48263; G1185010; -					
DR	MTM: 601459; -					
DR	PROSITE: PS01252; OPIOIDS_PRECURSOR: 1.					

OPPIOID PEPTIDE; NEUROTANSMITTER; NEUROPEPTIDE;
 KW CLEARANCE ON PAIR OF BASIC RESIDUES; SIGNAL.
 FT SIGNAL 1 19
 FT PEPTIDE 98 127 NEUROPEPTIDE 1 (PROBABLE).
 FT PEPTIDE 130 146 NOCICEPTIN (ORPHANIN FQ).
 FT PEPTIDE 149 165 NEUROPEPTIDE 2 (PROBABLE).
 SQ SEQUENCE 176 AA; 20295 MW; 8053843A CRC32;
 Query Match 100.0%; Score 53; DB 1; Length 176;
 Best Local Similarity 100.0%; Pred. No. 1.74e-03;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 169 TLHONGNV 176
 QY 1 TLHONGNV 8
 RESULT 2
 ID PNOG RAT STANDARD; PRT; 181 AA.
 AC 062923; 064162;
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE NOCICEPTIN PRECURSOR (ORPHANIN FQ) (PNOG) (ORL1 RECEPTOR AGONIST).
 GN PNOG.
 OS RATUS NORVEGICUS (RAT).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 OC EUTHERIA; RODENTIA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SPRAGUE-DAWLEY;
 RA NOTHACKER H. P., REINSCHEID R. K., MANSOUR A., HENNINGSEN R. A.,
 RA MONSMA F. J., JR., WATSON S. J., CIVELLI O.,
 RL SUBMITTED (FEB-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 96323281.
 RA MOLLEREAU C., SIMONS M.-J., SOULARUE P., LINERS F., VASSART G.,
 RA MEUNIER J.-C., PARMENTIER M.,
 RL PROC. NATL. ACAD. SCI. U.S.A. 93:8666-8670(1996).
 RN [3]
 RP SEQUENCE OF 13-181 FROM N.A., AND SEQUENCE OF 135-151.
 RC TISSUE-BRAIN;
 RX MEDLINE; 96013745.
 RA MEUNIER J.-C., MOLLEREAU C., TOLL L., SUAUDEAU C., MOISAND C.,
 RA ALVIERIE P., BUTOUR J.-L., GUILLET J.-C., FERRARA P.,
 RA MONSARRAT B., MAZARGUIL H., VASSART G., PARMENTIER M., COSTENTIN J.,
 RL NATURE 377:532-535(1995).
 RN [4]
 RP FUNCTION: NOCICEPTIN IS THE LIGAND OF THE OPIOID RECEPTOR-LIKE
 RECEPTOR (OPRL1). IT MAY ACT AS A TRANSMITTER IN THE BRAIN BY
 MODULATING NOCICEPTIVE AND LOCOMOTOR BEHAVIOR. MAY BE INVOLVED
 IN NEURONAL DIFFERENTIATION AND DEVELOPMENT.
 CC [1] SUBCELLULAR LOCATION: SECRETED INTO THE ENDOPLASMIC RETICULUM.
 CC [1] TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN THE SPINAL CORD
 AND BRAIN, BEING MORE ABUNDANT IN THE HYPOTHALAMUS AND STRIATUM.
 CC [1] ALSO FOUND IN SMALL AMOUNTS IN OVARY.
 CC [1] PTM: SPECIFIC ENZYMAIC CLEAVAGES AT PAIRED BASIC RESIDUES
 PROBABLY YIELD OTHER ACTIVE PEPTIDES BESIDES NOCICEPTIN.
 CC [1] PTM: THE N-TERMINAL CONTAINS 6 CONSERVED CYSTEINES THOUGHT TO
 BE INVOLVED IN DISULFIDE BONDING AND/OR PROCESSING.
 CC [1] SIMILARITY: BELONGS TO THE OPIOIDS NEUROPEPTIDES PRECURSORS
 FAMILY.
 DR EMBL; 048262; G1185012; -
 DR EMBL; X97375; E244795; -
 DR EMBL; S79730; G1172239; -
 DR PROSITE; PS01252; OPIOIDS_PRECURSOR; 1.
 KW OPIOID PEPTIDE; NEUROTANSMITTER; NEUROPEPTIDE;
 CLEAVAGE ON PAIR OF BASIC RESIDUES; SIGNAL; REPEAT.
 FT SIGNAL 1 19
 FT PEPTIDE 98 132 NEUROPEPTIDE 1 (PROBABLE).
 FT PEPTIDE 135 151 NOCICEPTIN (ORPHANIN FQ).
 FT PEPTIDE 154 170 NEUROPEPTIDE 2 (PROBABLE).
 FT DOMAIN 109 120 2 X 6 AA TANDEM REPEATS OF D-A-E-P-X-A.

REPEAT 109 114 1.
 FT REPEAT 115 120 2.
 SQ SEQUENCE 181 AA; 20478 MW; 1640366F CRC32;
 Query Match 100.0%; Score 53; DB 1; Length 181;
 Best Local Similarity 100.0%; Pred. No. 1.74e-03;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 174 TLHONGNV 181
 QY 1 TLHONGNV 8
 RESULT 3
 ID PNOG MOUSE STANDARD; PRT; 187 AA.
 AC 064387; 061105; 061938;
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE NOCICEPTIN PRECURSOR (ORPHANIN FQ) (PNOG) (N23K / N27K).
 GN PNOG OR NPNC1.
 OS MUS MUSCULUS (MOUSE).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 OC EUTHERIA; RODENTIA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE; 96216718.
 RA HOTTANT T., NISHI M., TAKESHIMA H., NUKADA T., SUGIMOTO T.,
 RA BIOCHEM. BIOPHYS. RES. COMMUN. 219:714-719(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 96106851.
 RA SAITO Y., MARYAMA K., SAITO T. C., KAWASHIMA S.,
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 217:539-545(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-129/SV;
 RX MEDLINE; 96323281.
 RA MOLLEREAU C., SIMONS M.-J., SOULARUE P., LINERS F., VASSART G.,
 RA MEUNIER J.-C., PARMENTIER M.,
 RL PROC. NATL. ACAD. SCI. U.S.A. 93:8666-8670(1996).
 RN [4]
 RP SEQUENCE FROM N.A. (LONG FORM).
 RX MEDLINE; 96279082.
 RA SAITO Y., MARYAMA K., KAWANO H., HAGINO-YAMAGISHI K.,
 RA KAWAMURA K., SAITO T. C., KAWASHIMA S.,
 RL J. BIOL. CHEM. 271:15615-15622(1996).
 RN [5]
 RP SEQUENCE OF 79-187 FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE; 96207555.
 RA PAN Y.-X., XU J., PASTERNAK G. W.,
 RL BIOCHEM. J. 315:11-13(1996).
 RN [6]
 RP FUNCTION: NOCICEPTIN IS THE LIGAND OF THE OPIOID RECEPTOR-LIKE
 RECEPTOR (OPRL1). IT MAY ACT AS A TRANSMITTER IN THE BRAIN BY
 MODULATING NOCICEPTIVE AND LOCOMOTOR BEHAVIOR. MAY BE INVOLVED
 IN NEURONAL DIFFERENTIATION AND DEVELOPMENT. WHEN ADMINISTERED
 INTRACEREBROVENTRICULARLY, NOCICEPTIN INDUCES HYPERALGESIA AND
 DECREASES LOCOMOTOR ACTIVITY.
 CC [1] SUBCELLULAR LOCATION: SECRETED INTO THE ENDOPLASMIC RETICULUM.
 CC [1] TISSUE SPECIFICITY: BRAIN AND SPINAL CORD. LOW LEVELS IN
 KIDNEY AND SPLEEN.
 CC [1] DEVELOPMENTAL STAGE: IN EMBRYONIC BRAIN, FIRST DETECTED AT DAY 14
 AND IN POSTNATAL BRAIN, LEVELS INCREASE IN DAY 1 AND DAY 18.
 CC [1] LEVELS DECREASE SIGNIFICANTLY IN ADULTS.
 CC [1] PTM: SPECIFIC ENZYMAIC CLEAVAGES AT PAIRED BASIC RESIDUES
 PROBABLY YIELD OTHER ACTIVE PEPTIDES BESIDES NOCICEPTIN.
 CC [1] PTM: THE N-TERMINAL CONTAINS 6 CONSERVED CYSTEINES THOUGHT TO
 BE INVOLVED IN DISULFIDE BONDING AND/OR PROCESSING.
 CC [1] ALTERNATIVE PRODUCTS: A LONGER FORM, N27K IS PROBABLY PRODUCED
 BY ALTERNATIVE SPLICING.
 CC [1] SIMILARITY: BELONGS TO THE OPIOIDS NEUROPEPTIDES PRECURSORS

CC FAMILY
 DR EMBL: D82866; D1012281; -
 DR EMBL: D50056; D1009395; -
 DR EMBL: X97373; E244874; -
 DR EMBL: X97371; E244782; -
 DR EMBL: X97372; E244782; JOINED.
 DR EMBL: U44027; G1335870; -
 DR EMBL: D50055; G1304162; -
 DR MGI: MGI:105308; NPNC1.
 DR PROSITE: PS0152; OPIOIDS_PRECURSOR; 1.
 DR OPIOID PEPTIDE: NEUROTRANSMITTER; NEUROPEPTIDE; ALTERNATIVE SPLICING;
 KM CLEAVAGE ON PAIR OF BASIC RESIDUES; SIGNAL; REPEAT.
 FT SIGNAL 1 19
 FT PEPTIDE 96 138
 FT PEPTIDE 141 157
 FT PEPTIDE 160 176
 FT DOMAIN 109 126
 FT REPEAT 109 114
 FT REPEAT 115 120
 FT REPEAT 121 126
 FT VARSPLIC 186 187
 FT CONFLICT 80 80
 FT SEQUENCE 187 AA; 20884 MW; F810ABP0 CRC32;
 SQ
 Query Match 100.0%; Score 53; DB 1; Length 187;
 Best Local Similarity 100.0%; Pred. No. 1.74e-03;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 180 TLHONGNV 187
 |||||
 QY 1 TLHONGNV 8

RESULT 4
 ID UBA2 WHEAT STANDARD; PRT: 1051 AA.
 AC P31251;
 DT 01-JUL-1993 (REL. 26, CREATED)
 DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE UBIQUITIN-ACTIVATING ENZYME E1 2.
 GN UBA2.
 OS TRITICUM AESTIVUM (WHEAT).
 CC EURARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; MONOCOTYLEDONEAE;
 CC CYPERALES; GRAMINEAE.
 CC [1]
 RP SEQUENCE FROM N.A.
 RN MEDLINE: 92340519.
 RA HATFIELD P.M., VIERSTRA R.D.;
 RL J. BIOL. CHEM. 267:14799-14803(1992).
 CC -1- FUNCTION: ACTIVATES UBIQUITIN BY FIRST ADENYLYLATING WITH ATP ITS
 CARBOXY-TERMINAL GLYCINE RESIDUE AND THEREAFTER LINKING THIS
 RESIDUE TO THE SIDE CHAIN OF A CYSTEINE RESIDUE IN E1, YIELDING
 AN UBIQUITIN-E1 THIOLESTER AND FREE AMP
 CC -1- PATHWAY: FIRST STEP IN UBIQUITIN CONJUGATION.
 CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
 CC -1- THERE ARE TWO ACTIVE SITES WITHIN THE E1 MOLECULE, ALLOWING IT TO
 ACCOMMODATE TWO UBIQUITIN MOIETIES AT A TIME, WITH A NEW UBIQUITIN
 FORMING AN ADENYLYL INTERMEDIATE AS THE PREVIOUS ONE IS
 TRANSFERRED TO THE THIOL SITE
 CC -1- THERE ARE MULTIPLE GENES ENCODING E1 IN WHEAT.
 CC -1- SIMILARITY: TO E1 IN OTHER SPECIES.
 CC -1- SIMILARITY: THE N-TERMINAL SHOWS SIMILARITY TO E. COLI CHLN.
 DR EMBL: M90663; G170684;
 DR PROSITE: PS00836; UBIQUITIN_ACTIVAT_1; 1.
 DR PROSITE: PS00865; UBIQUITIN_ACTIVAT_2; 1.
 KW UBIQUITIN CONJUGATION; LIGASE; MULTIGENE FAMILY; REPEAT.
 FT ACT SITE 626 626
 FT DOMAIN 56 605
 FT REPEAT 56 194
 FT REPEAT 453 605
 FT SEQUENCE 1051 AA; 116825 MW; F1376A64 CRC32;

Query Match 84.9%; Score 45; DB 1; Length 1051;
 Best Local Similarity 75.0%; Pred. No. 3.94e-01;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 93 TLHDDGNV 100
 |||||
 QY 1 TLHONGNV 8

RESULT 5
 ID UBA1 WHEAT STANDARD; PRT: 1051 AA.
 AC P20973;
 DT 01-FEB-1991 (REL. 17, CREATED)
 DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE UBIQUITIN-ACTIVATING ENZYME E1 1.
 GN UBA1.
 OS TRITICUM AESTIVUM (WHEAT).
 CC EURARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; MONOCOTYLEDONEAE;
 CC CYPERALES; GRAMINEAE.
 CC [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RN STRAIN-CV, AUGUSTA;
 RX MEDLINE: 90368797.
 RA HATFIELD P.M., CALLIS J., VIERSTRA R.D.;
 RL J. BIOL. CHEM. 265:15813-15817(1990).
 CC [2]
 RP MUTAGENESIS OF CYS, AND ACTIVE SITE.
 RN MEDLINE: 92340519.
 RA HATFIELD P.M., VIERSTRA R.D.;
 RL J. BIOL. CHEM. 267:14799-14803(1992).
 CC -1- FUNCTION: ACTIVATES UBIQUITIN BY FIRST ADENYLYLATING WITH ATP ITS
 CARBOXY-TERMINAL GLYCINE RESIDUE AND THEREAFTER LINKING THIS
 RESIDUE TO THE SIDE CHAIN OF A CYSTEINE RESIDUE IN E1, YIELDING
 AN UBIQUITIN-E1 THIOLESTER AND FREE AMP
 CC -1- PATHWAY: FIRST STEP IN UBIQUITIN CONJUGATION.
 CC -1- SUBUNIT: MONOMER.
 CC -1- PTM: THE N-TERMINUS IS BLOCKED.
 CC -1- THERE ARE TWO ACTIVE SITES WITHIN THE E1 MOLECULE, ALLOWING IT TO
 ACCOMMODATE TWO UBIQUITIN MOIETIES AT A TIME, WITH A NEW UBIQUITIN
 FORMING AN ADENYLYL INTERMEDIATE AS THE PREVIOUS ONE IS
 TRANSFERRED TO THE THIOL SITE.
 CC -1- THERE ARE MULTIPLE GENES ENCODING E1 IN WHEAT.
 CC -1- SIMILARITY: TO E1 IN OTHER SPECIES.
 CC -1- SIMILARITY: THE N-TERMINAL SHOWS SIMILARITY TO E. COLI CHLN.
 DR EMBL: M55604; G170780; -
 DR PIR: A38373; A38373.
 DR PIR: A42873; A42873.
 DR PROSITE: PS00536; UBIQUITIN_ACTIVAT_1; 1.
 DR PROSITE: PS00865; UBIQUITIN_ACTIVAT_2; 1.
 KW UBIQUITIN CONJUGATION; LIGASE; MULTIGENE FAMILY; REPEAT.
 FT ACT SITE 626 626
 FT DOMAIN 56 605
 FT REPEAT 56 194
 FT REPEAT 453 605
 FT SEQUENCE 1051 AA; 117007 MW; 614F10DB CRC32;

Query Match 84.9%; Score 45; DB 1; Length 1051;
 Best Local Similarity 75.0%; Pred. No. 3.94e-01;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 93 TLHDDGNV 100
 |||||
 QY 1 TLHONGNV 8

RESULT 6
 ID YEAP_ECOLI STANDARD; PRT: 384 AA.
 AC P76245; P94742; P97193;
 DT 15-JUL-1998 (REL. 36, CREATED)
 DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL 43.6 KD PROTEIN IN GAP-A-RND INTERGENIC REGION.

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 172 LHONGOV 178

QY 2 LHONGOV 8

RESULT 9 STANDARD: PRT: 261 AA.

AC P75713; P77128;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 28.7 KD PROTEIN IN GIP-PDRA INTERGENIC REGION.
GN YLBA OR GLXB6.
OS ESCHERICHIA COLI.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC ENTEROBACTERIACEAE.

RC STRAIN-K12 / MG1655;
RA BLATTNER F.R., PLUNKETT G. III, MAYHEW G.F., PERNA N.T., GLASNER F.D.;
RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.

RA ROBERTS D., ALLEN E., ARAUJO R., APARICIO A., CHUNG E., DAVIS K.,
RA DUNCAN M., FEDERSPIEL N., HYMAN R., KALMAN S., KOMP C., KUROI O.,
RA LEW H., LIN D., NAMATH A., OENNER P., SCHRAM S., DAVIS R.W.;
RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.

RC STRAIN-ECL1;
RA OBRADORS N., CUSA E., BALDOMA L., BADIA J., AGUILAR J.;
RL SUBMITTED (FEB-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AE000157; G1786725;
DR EMBL: U82664; G1773195;
DR ECGENE: EG13622; YLBA.
KW HYPOTHETICAL PROTEIN.
FT CONFLICT 204 204 A -> T (IN REF. 2).
FT CONFLICT 225 261 GDYFMGYSLQAGYGVGRGEAFSYISKDNRVEI ->
FT FT RRLHLYGRLEFTGLWLCRAW (IN REF. 2).
SQ SEQUENCE 261 AA; 28730 MW; C156E345 CRC32;

Query Match 77.4%; Score 41; DB 1; Length 261;
Best Local Similarity 100.0%; Pred. No. 4.75e+00;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 65 TLHONG 70

QY 1 TLHONG 6

RESULT 10 STANDARD: PRT: 508 AA.

AC P50264;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE FMS1 PROTEIN.

GN FMS1 OR YMR020W OR YMR711.09.
OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCYCETES.

RN [1]
RP SEQUENCE FROM N.A.
RA POUSETT D., MARCIREAU M., KARST F.;
RL SUBMITTED (DEC-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.

RC STAIN-S288C / AB972;
RA LYE G., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;
RL SUBMITTED (MAY-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- FUNCTION: MULTICOPY SUPPRESSOR OF PENICILLIN RESISTANCE.
CC -1- SIMILARITY: TO C.ALBICANS CBP1.

DR EMBL: X81848; E119324; -;
DR EMBL: Z49211; G798930; -;
DR SGD: L0003091; FMS1.
SQ SEQUENCE 508 AA; 57805 MW; 647D1742 CRC32;

Query Match 77.4%; Score 41; DB 1; Length 508;
Best Local Similarity 100.0%; Pred. No. 4.75e+00;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 26 TLHONG 31

QY 1 TLHONG 6

RESULT 11 STANDARD: PRT: 726 AA.

AC P25341;
DT 01-MAY-1992 (REL. 22, CREATED)
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE PROBABLE SERINE/THREONINE-PROTEIN KINASE KIN82 (EC 2.7.1.-).
GN KIN82 OR YCR91W OR YCR1153.
OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCYCETES.

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92254506.
RA WILSON C., BERGANTINO E., LANFRANCHI G., VALLE G., CARIGNANI G.,
RA FRONTALI L.;
RL YEAST 8.71-77(1992).

CC -1- SIMILARITY: TO S.POMBE SPAC48.05.
CC -1- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF SER/THR-
CC PROTEIN KINASES.
DR EMBL: X59720; E264713; -;
DR PIR: S22258; S22258.
DR HSSP: P05132; ICTP.
DR SGD: L0000906; KIN82.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
KW TRANSFERASE; SERINE/THREONINE-PROTEIN KINASE.
FT DOMAIN 324 602
FT NP-BIND 330 338 ATP (BY SIMILARITY).
FT BINDING 353 353 ATP (BY SIMILARITY).
FT ACT-SITE 449 449 BY SIMILARITY.
SQ SEQUENCE 726 AA; 82122 MW; A80B175B CRC32;

Query Match 77.4%; Score 41; DB 1; Length 726;
Best Local Similarity 71.4%; Pred. No. 4.75e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 457 LHONGHV 463

QY 2 LHONGHV 8

RESULT 12 STANDARD: PRT: 863 AA.

AC P32578;
DT 01-OCT-1993 (REL. 27, CREATED)
DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE SIP1 PROTEIN.

GN SIP1 OR YDR422C OR D9461.11.
OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCYCETES.

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92358217.
RA YANG X., HUBBARD E.J.A., CARLSON M.;
RL SCIENCE 257:680-682(1992).
RN [2]
RP SEQUENCE FROM N.A.

RA DIETRICH F.S., MULLIGAN J., ALLEN E., ARAUJO R., AVILES E.,
 RA BENO A., CARPENTER J., CHEN E., CHERRY J.M., CHONG E., DUNCAN M.,
 RA HOLWICK-SMITH S., HYMAN R., KOMP C., LASHKARI D., LEW H., LIN D.,
 RA MOSELEY D., NARAHARA K., NAMATH A., OEFNER P., OH C., PETEL F.X.,
 RA ROBERTS D., SCHRAMM S., SCHROEDER M., SHOGERN T., SHROFF N.,
 RA WIRANT A., YELTON M., BOTSTEIN D., DAVIS R.W.;
 RL SUBMITTED (AUG-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -1- FUNCTION: NOT KNOWN; INTERACTS WITH THE SNE1 PROTEIN KINASE.
 CC COULD BE AN ADAPTOR THAT PROMOTE THE ACTIVITY OF SNE1 TOWARDS
 CC SPECIFIC TARGETS.
 CC -1- PPM: PHOSPHORYLATED BY SNE1 IN VITRO.
 CC -1- SIMILARITY: BELONGS TO THE 5'-AMP-ACTIVATED PROTEIN KINASE, BETA
 CC SUBUNIT FAMILY.
 DR EMBL: M90531; G172604; -;
 DR EMBL: U33007; G927721; -;
 DR PIR: S41984; S41984.
 DR SGD: L0001890; S1P1.
 KM PHOSPHORYLATION.
 SQ SEQUENCE 863 AA; 96258 MW; 29D4E3DD CRC32;

Query Match 77.4%; Score 41; DB 1; Length 863;
 Best Local Similarity 62.5%; Pred. No. 4.75e+00;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 754 TLHONGNV 761
 QY 1 TLHONGNV 8

RESULT 13
 ID YDHF_ECOLI STANDARD; PRT; 298 AA.
 AC P76187;
 DT 15-JUL-1998 (REL. 36, CREATED)
 DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL OXIDOREDUCTASE IN SODC-NEMA INTERGENIC REGION.
 GN YDHF.
 OS ESCHERICHIA COLI.
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
 OC ENTEROBACTERIACEAE.
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RA BLATTNER F.R., PLUNKETT G. III, MAYHEW G.F., PERNA N.T., GLASNER F.D.;
 RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -1- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.
 CC STRONG, TO B.SUBTILIS YCSN.
 DR EMBL: AE000260; G1787936; -;
 DR ECOGENE: EG13420; YDHF.
 KM HYPOTHETICAL PROTEIN: OXIDOREDUCTASE.
 FT ACT SITE 128
 SQ SEQUENCE 298 AA; 33626 MW; 22B2EE26 CRC32;

Query Match 75.5%; Score 40; DB 1; Length 298;
 Best Local Similarity 71.4%; Pred. No. 8.61e+00;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 146 LHOSGKV 152
 QY 2 LHONGNV 8

RESULT 14
 ID GYRB_HALSQ STANDARD; PRT; 639 AA.
 AC P21358;
 DT 01-MAY-1991 (REL. 18, CREATED)
 DT 01-MAY-1991 (REL. 18, LAST SEQUENCE UPDATE)
 DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)
 DE DNA GYRASE SUBUNIT B (EC 5.99.1.3).
 GN GYRB.
 OS HALOERAX SP. (STRAIN AA 2.2).
 OC ARCHAEABACTERIA; EURYARCHAEOTA; HALOBACTERIALES; HALOBACTERIACEAE.
 RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE: 91100352.
 RA HOLMES M.L., DYALLSMITH M.L.;
 RL J. BACTERIOL. 173:642-648(1991).
 CC -1- FUNCTION: DNA GYRASE NEGATIVELY SUPERCOILS CLOSED CIRCULAR DOUBLE-
 CC STRANDED DNA IN AN ATP-DEPENDENT MANNER AND ALSO CATALYZES THE
 CC INTERCONVERSION OF OTHER TOPOLOGICAL ISOMERS OF DOUBLE-STRANDED
 CC DNA RINGS, INCLUDING CATENANES AND KNOTTED RINGS.
 CC -1- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING
 CC OF DOUBLE-STRANDED DNA.
 CC -1- SUBUNIT: MADE UP OF TWO CHAINS. THE A CHAIN IS RESPONSIBLE FOR DNA
 CC BREAKAGE AND REJOINING, THE B CHAIN CATALYZES ATP HYDROLYSIS.
 CC -1- MOTATIONS IN DNA GYRASE RESULT IN NOVIOICIN RESISTANCE IN
 CC HALOPHILIC ARCHAEABACTERIA.
 DR EMBL: M38373; G149024; -;
 DR PIR: A39135; A39135.
 DR PROSITE: P500177; TOPOISOMERASE II; 1.
 KM TOPOISOMERASE, ISOMERASE, ATP-BINDING.
 SQ SEQUENCE 639 AA; 71138 MW; FFA41166 CRC32;

Query Match 75.5%; Score 40; DB 1; Length 639;
 Best Local Similarity 50.0%; Pred. No. 8.61e+00;
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 63 ALHEDGSV 70
 QY 1 TLHONGNV 8

RESULT 15
 ID SP22_BACME STANDARD; PRT; 147 AA.
 AC P35148;
 DT 01-FEB-1994 (REL. 28, CREATED)
 DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
 DE ANTI-SIGMA F FACTOR (STAGE II SPOULATION PROTEIN AB).
 GN SPOI1AB.
 OS BACILLUS MEGATERIUM.
 OC PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.
 RN (1)
 RP SEQUENCE FROM N.A.
 RC MEDLINE: 93003525.
 RA TAO Y.P., HUDSPETH D.S.S., VARY P.S.;
 RL BIOCHIMIE 74:695-704(1992).
 CC -1- FUNCTION: BINDS TO SIGMA F AND BLOCKS ITS ABILITY TO FORM AN RNA
 CC POLYMERASE HOLOENZYME (E-SIGMA F). PHOSPHORYLATES SPOI1AA ON A
 CC SERINE RESIDUE. THIS PHOSPHORYLATION MAY ENABLE SPOI1AA TO ACT AS
 CC AN ANTI-ANTI-SIGMA FACTOR THAT COUNTERACTS SPOI1AB AND THUS
 CC RELEASES SIGMA F FROM INHIBITION (BY SIMILARITY).
 CC -1- SIMILARITY: STRONG, TO RBSW.
 DR EMBL: X63757; G39653; -;
 DR PIR: S22178; S22178.
 DR PIR: B48402; B48402.
 KM SPOI1ATION; TRANSFERASE; KINASE.
 SQ SEQUENCE 147 AA; 16382 MW; CD51F1A5 CRC32;

Query Match 73.6%; Score 39; DB 1; Length 147;
 Best Local Similarity 62.5%; Pred. No. 1.54e+01;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 69 TLHEDGV 76
 QY 1 TLHONGNV 8

Search completed: Fri Apr 16 14:26:31 1999
 Job time : 8 secs.

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

```

MPSrch_gp protein - protein database search, using Smith-Waterman algorithm
Run on:      Fri Apr 16 14:26:48 1999;      MasPar time 4.55 Seconds
Tabular output not generated.      87.470 Million cell updates/sec

```

```
Title: >US-09-011-797-4
Description: (1-8) from US09011797.pep
Perfect Score: 53
Sequence: 1 TLHQNGNV 8
```

Scoring table: PAM 150
Gap 15

Searched: 165420 seqs, 49795644 residues

```
Post-processing: Minimum Match 0%
                  Listing first 45 summaries
```

Database:

spiremb16

1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_orig
9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclass
13:sp_vertebrate 14:sp_virus

Statistics: Mean 19.822; Variance 18.814; scale 1.054

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	46	86.8	176	6	063647	NOCICEPTIN/ORPHANIN FQ	4.57e+00
2	43	81.1	251	2	025087	CONSERVED RHETROPHETICAL	3.01e+00
3	43	81.1	636	2	P77993	DNA TOPISOMERASE II S	3.01e+00
4	43	81.1	2150	5	044131	C08G9.2 PROTEIN.	3.01e+00
5	42	79.2	221	5	Q22783	COSMID T25G12.	5.50e+00
6	41	77.4	221	2	054410	CSXB6.	9.95e+00
7	41	77.4	303	10	P92932	LECTIN RELATED PROTEIN	9.95e+00
8	41	77.4	351	2	069578	PUTATIVE LONG CHAIN FA	9.95e+00
9	41	77.4	507	13	057439	TU12.	9.95e+00
10	41	77.4	507	13	013020	AMINO ACID TRANSPORTER	9.95e+00
11	41	77.4	1275	14	041877	TEGUMENT PROTEIN/PEARA	9.95e+00
12	40	75.5	328	1	028652	CONSERVED RHETROPHETICAL	1.78e+00
13	40	75.5	609	5	016938	F26G5.1 PROTEIN.	1.78e+00
14	40	75.5	745	2	032114	YIRC PROTEIN.	1.78e+00
15	39	73.6	74	2	Q49718	B1549.F3.145.	3.15e+00
16	39	73.6	158	2	046245	NITROGENASE (FRAGMENT)	3.15e+00
17	39	73.6	192	2	007135	B1306.G3 PROTEIN.	3.15e+00
18	39	73.6	232	10	040590	THRP-1 PROTEIN.	3.15e+00
19	39	73.6	329	2	046084	NITROGENASE 3 SUBUNIT	3.15e+00
20	39	73.6	439	1	023985	NADH OXIDASE (NOKA-1).	3.15e+00

	21	39	73.6	453	13	042115	ARX HOMEOPROREIN.	3.15e+01
	22	39	73.6	623	3	042626	PROTEIN KINASE NRC-2,	3.15e+01
	23	39	73.6	720	3	006472	BETA-GUCAN SYNTHESIS-	3.15e+01
	24	39	73.6	765	5	021478	SIMILARITY OVER A SHOR	3.15e+01
	25	39	73.6	880	11	P97693	P105 COACTIVATOR.	3.15e+01
	26	39	73.6	885	4	Q13122	100 KDA COACTIVATOR.	3.15e+01
	27	39	72.6	890	3	060040	KID RNA PROTEIN.	3.15e+01
	28	39	73.6	1288	11	P97608	5-OXO-L-POLYNASE.	3.15e+01
	29	38	71.7	80	14	055781	NON-STRUCTURAL PROTEIN	5.50e+01
	30	38	71.7	238	1	027227	N5-METHYL-TETRAHYDROME	5.50e+01
	31	38	71.7	249	1	028475	CONSERVED HYPOTHEICAL	5.50e+01
	32	38	71.7	325	2	067552	HYPOTHEICAL 36.2 KD P	5.50e+01
	33	38	71.7	362	5	001493	SIMILARITY TO THE CAT	5.50e+01
	34	38	71.7	471	11	035162	MICROSOMAL STRESS 70 P	5.50e+01
	35	38	71.7	550	3	039725	HYPOTHEICAL 63.2 KD P	5.50e+01
	36	38	71.7	585	13	057366	RTSOX23.	5.50e+01
	37	38	71.7	601	10	023522	TRIACYLGLYCEROL LIPASE	5.50e+01
	38	38	71.7	636	11	035656	FLEISCHEN.	5.50e+01
	39	38	71.7	639	3	043113	ARYLSULFATASE.	5.50e+01
	40	38	71.7	663	14	087039	SPVCPZ, COMPLETE GENOM	5.50e+01
	41	38	71.7	669	11	054770	LENS FIBER CELL BEADED	5.50e+01
	42	38	71.7	718	2	034587	AMX PROTEIN.	5.50e+01
	43	38	71.7	754	11	061123	MATERIENL EMBRYONIC MES	5.50e+01
	44	38	71.7	1053	5	024310	POLYMERPROTEIN.	5.50e+01
	45	38	71.7	1337	2	P76465	F1337.	5.50e+01

RESULT	1	ALIGNMENTS
ID	062647	PRELIMINARY; PRT; 176 AA.
AC	062647;	
DT	01-AUG-1998 (TREMBLREL. 07, CREATED)	
DT	01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)	
DT	01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)	
DE	NOICEPTIN/ORPHANIN FQ PRECURSOR.	
OS	BOS TAURUS (BOVINE).	
OC	EUDARTOTA; METAFOCA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;	
CC	EUDARTOTA; ARTIODACTYLA.	
RN	[1]	
RA	SEQUENCE FROM N.A.	
RP	OXIDA-ASHTAKA E., MINAMI T., TACHIBANA S., YOSHIMURA Y.,	
RL	NISHIKUCHI Y., KIMURA T., ITO S.;	
DR	NATURE 392:286-289(1996).	
SR	EMBL; AB005251; D1026347; -	
SEQ	SEQUENCE 176 AA; 20255 MW; D2C23826 CRC32;	

Query Match	86.8%;	Score 46;	DB 6;	Length 176;
Best Local Similarity 100.0%;	Fred. No. 4.67e-01;			
Matches 7; Conservative	0; Mismatches 0; Indels 0;	Gaps 0;		

Db	169 TLHONGN 175
Cy	1 TLHONGN 7

RESULT	2
ID	025087
AC	025087;
DT	01-JAN-1998 (TREMBLREL. 05, CREATED)
DT	01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT	01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE	CONSERVED HYPOTHEICAL PROTEIN.
GN	HP0318.
OS	HELICOBACTER PYLORI (CAMPILOBACTER PYLORI).
OC	PROKARYOTA; GRACILICUTES; SCOTOBACTERIA;
CC	AEROBIC, MOTILE, HELICAL AND/OR VIBRIOID.
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN-26695;
RA	MEDLINE; 97394467.
TX	TOMB J.-F., WHITE O., KERLAVAGE A.R., CLAYTON R.A., SUTTON G.G.,
FL	FLEISCHMANN R.D., KETCHUM K.A., KLENK H.-P., GILL S., DOUGHERTY B.A.,


```

RA NELSON K., QUACKENBUSH J., ZHOU L., KIRKNESS E.F., PETERSON S.,
RA LOFTUS B., RICHARDSON D., DODSON R., KHALAK H.G., GLODER A.,
RA KEMNENY K., FITZGERALD L.M., LEE N., ADAMS M.D., HICKEY E.K.,
RA BERG D.E., GOCAYNE J.D., UTTERBACK T.R., PETERSON J.D., KELLEY J.M.,
RA COTTON M.D., WEIDMAN J.M., FUJII C., BOWMAN C., WATNEY L., WALLIN E.,
RA HAYES W.S., BORODOVSKIY M., KARP P.D., SMITH H.O., FRASER C.M.,
RA VENTER J.C.,
RL NATURE 368:539-547(1997).
DR EMBL: AE000550; G2313418; -.
DR TIGR: HP0318; -.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 251 AA; 28507 MW; F7119AC2 CRC32;

Query Match
Best Local Similarity 81.1%; Score 43; DB 2; Length 251;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 105 TLHNGHV 112
OY 1 TLHNGNV 8

RESULT 3
ID P77993; PRELIMINARY; PRT; 636 AA.
AC P77993;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DE 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE DNA TOPOISOMERASE II SUBUNIT B.
GN TOP2B.
OS THERMOTOGA MARITIMA.
OC PROKARYOTA; NOT YET CLASSIFIED.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MSB8.
RL MEDLINE: 97017137.
RA GUIPAUD O., LABEDAN B., FORTERRE P.;
RL GENE 174:121-128(1996).
CC -1- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING
OF DOUBLE-STRANDED DNA.
DR EMBL: U49693; G1622792; -.
DR PROSITE: PS00177; TOPOISOMERASE-II; 1.
DR PFAM: PF00204; DNA_TOPOISOIL.
KW ISOMERASE; TOPOISOMERASE; ATP-BINDING.
SQ SEQUENCE 636 AA; 72459 MW; D9915C88 CRC32;

Query Match
Best Local Similarity 81.1%; Score 43; DB 2; Length 636;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 60 TLHNGSV 67
OY 1 TLHNGNV 8

RESULT 4
ID 044131; PRELIMINARY; PRT; 2150 AA.
AC 044131;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DE 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE C0869.2 PROTEIN.
GN C0869.2.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; ACCELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2.
RL MEDLINE: 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,
RA BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAXTON M.,
RA DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,
RA HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,

```

```

RA KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,
RA MCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,
RA RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SWALDON N., SMITH A.,
RA SONNHAMMER E., STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K.,
RA VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,
RA WILKINSON-SPROAT J., WOHLDMAN P.;
RL NATURE 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2.
RA GEISEL C., STELLYES L.;
RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2.
RA WATERSTON R.;
RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AF036687; G2662541; -.
DR PROSITE: PS00280; BPTI_KUNITZ; 1.
DR PROSITE: PS00317; 4-DISULFIDE_CORE; 9.
KW SERINE PROTEASE INHIBITOR.
SQ SEQUENCE 2150 AA; 234130 MW; 8231D798 CRC32;

Query Match
Best Local Similarity 81.1%; Score 43; DB 5; Length 2150;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1641 LHONGNV 1647
OY 2 LHONGNV 8

RESULT 5
ID 022783; PRELIMINARY; PRT; 771 AA.
AC 022783;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE COSMID T25G12.
GN T25G12.6.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; ACCELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN [1]
RP SEQUENCE FROM N.A.
RL MEDLINE: 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,
RA BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAXTON M.,
RA DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,
RA HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,
RA KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,
RA MCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,
RA RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SWALDON N., SMITH A.,
RA SONNHAMMER E., STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K.,
RA VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,
RA WILKINSON-SPROAT J., WOHLDMAN P.;
RL NATURE 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2.
RA DU Z.;
RL SUBMITTED (JAN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC WATERSTON R.;
RL SUBMITTED (DEC-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: U43283; G1125834; -.
SQ SEQUENCE 771 AA; 83449 MW; 9B6C7ABA CRC32;

Query Match
Best Local Similarity 79.2%; Score 42; DB 5; Length 771;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 722 TLHNGKV 729
OY 1 TLHNGNV 8

```


OY 1 TLHONGNV 8

RESULT 6
ID 054410 PRELIMINARY; PRT: 221 AA.
AC 054410:
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE GLXB6.
GN GLXB6.
OS ESCHERICHIA COLI.
OC PROKARYOTA: GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC ENTEROBACTERIACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RA OBRAODORS N., CUSA E., BALDOMA L., BADIA J., AGUILAR J.;
RL SUBMITTED (FEB-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: U89279; G2735241; -
SO SEQUENCE 221 AA; 24306 MW; C91F86AA CRC32;

Query Match 77.4%; Score 41; DB 2; Length 221;
Best Local Similarity 100.0%; Pred. No. 9.95e+00;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 65 TLHONG 70
|||
OY 1 TLHONG 6

RESULT 7
ID P92932 PRELIMINARY; PRT: 303 AA.
AC P92932:
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE LECTIN RELATED PROTEIN PRECURSOR (FRAGMENT).
OS ALLIUM SATIVUM (CARLIC).
OC EUKARYOTA: PLANTA: EMBRYOPHYTA: ANGIOSPERMAE: MONOCOTYLEDONEAE;
OC LILIALES; LILIACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-ROOTS;
RX MEDLINE: 97188556.
RA SNETS K., VAN DAME E., VERHAERT P., BARRE A., VAN LEUVEN F.,
RA PEDMANS W.J.;
RL PLANT MOL. BIOL. 33:223-234(1997).
DR EMBL: U58949; G1840051; -
KW SIGNAL; LECTIN.
FT NON_TER 1
FT SIGNAL <1 12 POTENTIAL,
FT CHAIN 13 303 LECTIN RELATED PROTEIN.
SQ SEQUENCE 303 AA; 32749 MW; 6D218BC0 CRC32;

Query Match 77.4%; Score 41; DB 10; Length 303;
Best Local Similarity 71.4%; Pred. No. 9.95e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 242 LQDQGNV 248
|||
OY 2 LQDQGNV 8

RESULT 8
ID 069578 PRELIMINARY; PRT: 351 AA.
AC 069578:
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE PUTATIVE LONG CHAIN FATTY ACID-COA LIGASE.
GN MLCB268.30.
OS MYCOBACTERIUM LEPRAE.
OC PROKARYOTA: FIRMICUTES; ACTINOMYCETALES; MYCOBACTERIACEAE.

RN [1]
RP SEQUENCE FROM N.A.
RA BROWN D., CHURCHER C.M.;
RL SUBMITTED (APR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RA PARKHILL J., BARRELL B.G., RAJANDREAN M.A.;
RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RA EIGMEIER K., HONORE N., WOODS S.A., CAUDRON B., COLE S.T.;
RL MOL. MICROBIOL. 7:197-206(1993).
DR EMBL: AL022602; F1287820; -
KW LIGASE.
SO SEQUENCE 351 AA; 37624 MW; 76BC96A3 CRC32;

Query Match 77.4%; Score 41; DB 2; Length 351;
Best Local Similarity 75.0%; Pred. No. 9.95e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

DB 136 TLNDAQNV 143
|||
OY 1 TLHONGNV 8

RESULT 9
ID 057439 PRELIMINARY; PRT: 507 AA.
AC 057439:
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE IUI2.
OS XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC EUKARYOTA: METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AMPHIBIA; ANURA.
RN [1]
RP SEQUENCE FROM N.A.
RA SHI Y.;
RL SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AF019066; G2731774; -
SO SEQUENCE 507 AA; 55595 MW; 4581F63F CRC32;

Query Match 77.4%; Score 41; DB 13; Length 507;
Best Local Similarity 100.0%; Pred. No. 9.95e+00;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 28 LHONGN 33
|||
OY 2 LHONGN 7

RESULT 10
ID 013020 PRELIMINARY; PRT: 507 AA.
AC 013020:
DT 01-JUN-1997 (TREMBLREL. 04, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE AMINO ACID TRANSPORTER CHAIN ASUR4.
GN ASUR4.
OS XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC EUKARYOTA: METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AMPHIBIA; ANURA.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-KIDNEY;
RA SPINDLER B., MASTROBERARDINO L., CUSTER M., VERREY F.;
RL PELIGERS ARCH. 434:323-331(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-KIDNEY;
RA VERREY F.;
RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE-KIDNEY;

RA VERREY F.;
 RL SUBMITTED (APR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC TISSUE-KIDNEY;
 RA SHI X.;
 RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: Y12716; E1289893;
 SQ SEQUENCE 507 AA; 55529 MW; CEF914A3 CRC32;
 Query Match 77.4%; Score 41; DB 13; Length 507;
 Best Local Similarity 100.0%; Pred. No. 9.95e+00;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 28 LHONGN 33
 111111
 2 LHONGN 7

RESULT 11
 ID 041977 PRELIMINARY; PRT; 1275 AA.
 AC 041977;
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
 DE TEGUMENT PROTEIN/EGARAT.
 GN GAMMAHV ORF75B.
 OS MURINE HERPESVIRUS 68.
 CC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE; GAMMAHERPESVIRINAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-WUMS;
 RX MEDLINE: 97366649.
 RA VIRGIN H.W. IV, LATREILLE P., WANSLEY P., HALLSWORTH K., WECK K.E.,
 RA DAL CANTO A.J., SPECK S.H.;
 RL J. VIROL. 71:5894-5904(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-WUMS;
 RA LATREILLE P., WANSLEY P., WATERSTON R.H.;
 RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: U97553; G2318001;
 SQ SEQUENCE 1275 AA; 141953 MW; 518C97E0 CRC32;
 Query Match 77.4%; Score 41; DB 14; Length 1275;
 Best Local Similarity 62.5%; Pred. No. 9.95e+00;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 Db 1057 TLHONGV 1064
 111111
 1 TLHONGV 8

RESULT 12
 ID 028652 PRELIMINARY; PRT; 328 AA.
 AC 028652;
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
 DE 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
 GN CONSERVED HYPOTHETICAL PROTEIN.
 OS ARCHAEoglobus FULGIDUS.
 CC ARCHAEOBACTERIA; EURYARCHAEOTA; ARCHAEoglobales; ARCHAEoglobaceae.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-VC-16 / DSM 4304 / ATCC 49558;
 RX MEDLINE: 98049343.
 RA KLEIN H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,
 RA KITCHOW K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,
 RA RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KYRPIES N.C.,
 RA FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,
 RA KRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,
 RA PETERSON S., REICH C.I., MCNEIL L.R., BADGER J.H., GLODER A., ZHOU L.,

RA OVERBECK R., GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTERBACK T.,
 RA COTTON M.D., SPRIGGS T., ARTIACH P., KAINE B.P., SYKES S.M.,
 RA SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
 RA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,
 RA VENTER J.C.;
 RL NATURE 390:364-370(1997).
 DR EMBL: AE000990; G2648933;
 DR TIGR: AF1621;
 KW HYPOTHETICAL PROTEIN.
 SQ SEQUENCE 328 AA; 35412 MW; F803A6BD CRC32;
 Query Match 75.5%; Score 40; DB 1; Length 328;
 Best Local Similarity 75.0%; Pred. No. 1.78e+01;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Db 198 TLHONGV 205
 111111
 1 TLHONGV 8

RESULT 13
 ID 016938 PRELIMINARY; PRT; 609 AA.
 AC 016938;
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
 DE F26G5.1 PROTEIN.
 DE F26G5.1 PROTEIN.
 GN F26G5.1
 OS CAENORHABDITIS ELEGANS.
 CC EURYARCHAEOTA; METAZOA; ACCELLOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE: 94150718.
 RA WILSON R., AINSWORTH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,
 RA BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAYTON M.,
 RA DEAR S., DU Z., DUBBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,
 RA HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,
 RA KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,
 RA MCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,
 RA RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SALLDON N., SMITH A.,
 RA SONNHAMMER E., STADEN R., SULLSTON J., THIERRY-MEG J., THOMAS K.,
 RA VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,
 RA WILKINSON-SPROAT J., WOHLDMAN P.;
 RL NATURE 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA SAMMONS L., WOHLDMAN P., BECK C.;
 RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: AF022974; G2384841;
 SQ SEQUENCE 609 AA; 68849 MW; 804D9451 CRC32;

Query Match 75.5%; Score 40; DB 5; Length 609;
 Best Local Similarity 62.5%; Pred. No. 1.78e+01;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 Db 506 TLHONGV 513
 111111
 1 TLHONGV 8
 RESULT 14
 ID 032144 PRELIMINARY; PRT; 745 AA.
 AC 032144;
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
 DE 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
 GN YURC PROTEIN.
 OS BACILLUS SUBTILIS.
 CC PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RA KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONI G.,
 RA AAEVERO V., BERTERO M.G., BESSIERES P., BOLOTIN A., BORCHERT S.,
 RA BORRIS R., BOUSIER L., BRANS A., BRAUN M., BRIGHELL S.C., BRON S.,
 RA BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,
 RA CHOI S.K., CODANI J.J., CONNERTON I.F., CUMINGS N.J., DANIEL R.A.,
 RA DERIZOT F., DEVINE K.M., DUSTERHOFF A., EHRLICH S.D., EMERSON P.T.,
 RA ENITAN K.D., ERRINGTON J., FABRET C., FERRARI E., FOULGER D., FRITZ C.,
 RA FUJITA M., FUJITA Y., FUMA S., GALIZZI A., GALLERON N., GHIM S.Y.,
 RA GLASER P., GOFFEAU A., GOLIGHTLY E.J., GRANDI G., GUSEPPI G.,
 RA GUY B.J., HAGA K., HAIECH J., HARWOOD C.R., HENAUT A., HILBERT H.,
 RA HOLSAPEL S., HOSONO S., HULLO M.F., ITRAYA M., JONES L., JOHNS B.,
 RA KARAMATA D., KASAHARA Y., KLARER-BLANCHARD M., KLEIN C., KOBAYASHI Y.,
 RA KOETTER P., KONIGSSTEIN G., KROGH S., KUMANO M., KURITA K., LAPIDUS A.,
 RA LARDINOIS S., LAUBER J., LAZAREVIC V., LEE S.M., LEVINE A., LIU H.,
 RA MASUDA S., MAUEL C., MEDIGUE C., MEDINA N., MELLAO R.P., MIZUNO M.,
 RA MOESTL D., NAKAI S., NOBACK M., NOONE D., O'REILLY M., OGAWA K.,
 RA OGIMAWA A., OUDEGA B., PARK S.H., PARO V., POHL T.M., PORTELELE D.,
 RA PORMOLIK S., PRESCOTT A.M., PRESCAN E., PUTIC P., PURNELLE B.,
 RA RAPOPORT G., REY M., REYNOLDS S., RIEGER M., RIVOLTA C., ROCHA E.,
 RA ROCHE B., ROSE M., SADAIE Y., SATO T., SCAMIAN E., SCHLEICH S.,
 RA SCHROETER R., SCOFFONE F., SEKIGUCHI J., SEKONSKA A., SEROR S.J.,
 RA SEROR P., SHIN B.S., SOLDO B., SOROKIN A., TACCONI E., TAKAGI T.,
 RA TAKAHASHI H., TAKEMARU K., TAKEUCHI M., TANAKOSHI A., TANAKA T.,
 RA TERPSTRA P., TOGNONI A., TOSATO V., UCHITAMA S., VANDENBOL M.,
 RA VANNIER F., VASSAROTTI A., VIARI A., WABUTT R., WEDLER E., WEDLER H.,
 RA WEITZENEGGER T., WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K.,
 RA YASUMOTO K., YATA K., YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E.,
 RA YOSHIKAWA H., DANCHIN A.;
 RL NATURE 390:249-256(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RA KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;
 RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: 299120: E1184327: -
 SQ SEQUENCE 745 AA: 80433 MW: D96C4EAB CRC32:

Query Match 75.5%; Score 40; DB 2; Length 745;
 Best Local Similarity 57.1%; Pred. No. 1.78e+01;
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 140 LHEDGNT 146
 QY 2 LHONGNV 8

RESULT 15
 ID Q49718 PRELIMINARY; PRT; 74 AA.
 AC Q49718;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
 DE B1549_F3_145;
 OS MYCOBACTERIUM LEPRAE;
 OC PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; MYCOBACTERIACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RL SUBMITTED (MAR-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RL SUBMITTED (NOV-1993) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [3]
 RP SEQUENCE FROM N.A.
 RL SMITH D.R.;
 RL SUBMITTED (JAN-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [4]
 RP SEQUENCE FROM N.A.
 RL ROBINSON K.;
 RL SUBMITTED (SEP-1994) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL: U00014: G466929: -
 SQ SEQUENCE 74 AA: 7972 MW: 56A2486D CRC32:
 Query Match 73.6%; Score 39; DB 2; Length 74;
 Best Local Similarity 71.4%; Pred. No. 3.15e+01;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 48 TLHEHGN 54
 QY 1 TLHONGN 7

Search completed: Fri Apr 16 14:27:20 1999
 Job time : 32 secs.

THIS PAGE BLANK (USPTO)

 [W] [O] [R] [L] (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPsrch_tpn n.a. - n.a. Smith-Waterman search, using a protein query
 which has been backtranslated into n.a. using IUPAC symbols

Run on: Tue Apr 27 10:33:06 1999: MasPar time 99.54 Seconds

Tabular output not generated. 431.758 Million cell updates/sec

Title: >US-09-011-797-4
 Description: (1-8) from US09011797.pep
 Perfect Score: 100
 N.A. Sequence: 1 ACNTYNCAYCARAAYGNAAYGTN 24
 Comp: TGNRANGTRGTYTRCMTTCRACN

Scoring table: TABLE backtranslated
 Gap 40

Mmatch STD: Dbase 0; Query 0

Searched: 2275026 segs, 89538244 bases x 2

Post-processing: Minimum Match 08
 Listing first 45 summaries

Database: emb1-est56
 1:em-est1 2:em-gss1 3:em-gss2 4:em-gss3
 genbank-est109
 5:gb-est1 6:gb-est10 7:gb-est11 8:gb-est12 9:gb-est13
 10:gb-est14 11:gb-est15 12:gb-est16 13:gb-est17
 14:gb-est18 15:gb-est19 16:gb-est20 17:gb-est21
 18:gb-est22 19:gb-est23 20:gb-est24 21:gb-est25 22:gb-est26
 23:gb-est27 24:gb-est28 25:gb-est29 26:gb-gss1 27:gb-gss2
 28:gb-gss3 29:gb-gss4

Statistics: Mean 40.359; Variance 39.952; scale 1.010

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
C	1	95	95.0	402	15	AI007467	L0-370M13R Ice plant L	3.12e-04
C	2	95	95.0	431	14	AA962868	L30-398T3 Ice plant La	3.12e-04
C	3	95	95.0	471	14	AA962918	L30-449T3 Ice plant La	3.12e-04
C	4	95	95.0	472	14	AA887369	L30-342T3 Ice plant La	3.12e-04
C	5	95	95.0	482	14	AA962885	L30-415T3 Ice plant La	3.12e-04
C	6	95	95.0	486	14	AA962900	L30-430T3 Ice plant La	3.12e-04
C	7	95	95.0	487	14	AA962902	L30-433T3 Ice plant La	3.12e-04
C	8	85	85.0	209	8	AA213399	zr1910.r1 NCI-CGAP_GC	9.90e-02
C	9	85	85.0	236	16	N41298	yw68604.r1 Homo sapien	9.90e-02
C	10	85	85.0	285	16	T35252	EST81954 Homo sapien	9.90e-02
C	11	85	85.0	290	19	H92421	YS86603.r1 Homo sapien	9.90e-02
C	12	85	85.0	293	21	T19059	B060055 Testis 2 Homo	9.90e-02

C	13	85	85.0	305	16	T34361	EST66604 Homo sapien	9.90e-02
C	14	85	85.0	371	26	B45681	HS-1062-A2-A06-MF-ab1	9.90e-02
C	15	85	85.0	377	15	C83092	Oryctolagus cuniculus	9.90e-02
C	16	85	85.0	377	15	C83225	Oryctolagus cuniculus	9.90e-02
C	17	85	85.0	383	24	AA295789	EST100992 Pancreas tum	9.90e-02
C	18	85	85.0	390	5	T99944	ye72e03.r1 Homo sapien	9.90e-02
C	19	85	85.0	416	5	T75533	ye63e02.r1 Homo sapien	9.90e-02
C	20	85	85.0	420	5	T99283	ye63f01.r1 Homo sapien	9.90e-02
C	21	85	85.0	443	14	AA837379	od41c11.s1 NCI-CGAP_GC	9.90e-02
C	22	85	85.0	482	24	AA13767	EST185694 Colon carcin	9.90e-02
C	23	85	85.0	504	6	AA195688	zr32h11.r1 Soares NIM	9.90e-02
C	24	80	80.0	169	18	AA159267	vz8e01.r1 Soares 2NM	1.46e+00
C	25	80	80.0	307	15	AA96374	OS06c05.s1 NCI-CGAP_Lu	1.46e+00
C	26	80	80.0	310	19	HSC2SE022	H. sapiens partial cDN	1.46e+00
C	27	80	80.0	321	9	AA597789	29856 Lambda-PRL2 Arab	1.46e+00
C	28	80	80.0	334	16	H00125	Y169405.r1 Homo sapien	1.46e+00
C	29	80	80.0	351	12	AA731954	nw64h07.s1 NCI-CGAP_GC	1.46e+00
C	30	80	80.0	364	16	HSC16A051	H. sapiens partial cDN	1.46e+00
C	31	80	80.0	390	13	AA810584	oa85h10.s1 NCI-CGAP_GC	1.46e+00
C	32	80	80.0	406	7	AA447072	z886002.r1 Soares tota	1.46e+00
C	33	80	80.0	407	26	B63132	CIT978SK-A-657C7.rp CI	1.46e+00
C	34	80	80.0	413	13	AA180411	z532B03.r1 Homo sapien	1.46e+00
C	35	80	80.0	414	20	N99782	Y199f01.r1 Homo sapien	1.46e+00
C	36	80	80.0	421	16	H01380	HS-3218.B1.B03.MR CIT	1.46e+00
C	37	80	80.0	429	28	AO181628	HS-1056-B1-C03-MF-ab1	1.46e+00
C	38	80	80.0	440	26	B42802	zs48c08.s1 NCI-CGAP_GC	1.46e+00
C	39	80	80.0	467	8	AA285125	Y11b10.r1 Homo sapien	1.46e+00
C	40	80	80.0	484	16	R62848	aag3e09.s1 Homo sapien	1.46e+00
C	41	80	80.0	489	13	AA464930	v97Bb09.r1 Barstead MP	1.46e+00
C	42	80	80.0	491	7	AA473537	aj44b12.s1 Soares test	1.46e+00
C	43	80	80.0	545	12	AA813019	CIT-HSP-236108.TR CIT	1.46e+00
C	44	80	80.0	562	28	AO111189	T26N18-T7 TAMU Arabido	1.46e+00
C	45	80	80.0	1005	26	B11855		1.46e+00

ALIGNMENTS

RESULT 1
 LOCUS AI007467 402 bp mRNA
 DEFINITION L0-370M13R Ice plant Lambda Uni-Zap XR expression library, 0 hours
 NACL treatment Mesembryanthemum crystallinum cDNA clone L0-370 5'
 similar to 40S RIBOSOMAL PROTEIN S4, mRNA sequence.

ACCESSION AI007467
 NID 93216977
 KEYWORDS EST.

SOURCE Mesembryanthemum crystallinum
 ORGANISM Common ice plant.

REFERENCE 1 (bases 1 to 402)
 Cushman, J.C.
 AUTHORs Eukaryote; Viridiplantae; Charophyta/Embryophyta group;
 Embryophyta; Tracheophyta; seed plants; Magnoliophyta;
 eudicotyledons; Caryophyllidae; Caryophyllaceae;
 Mesembryanthemum.

TITLE An expressed sequence tag database for the common ice plant,
 Mesembryanthemum crystallinum
 JOURNAL Unpublished (1997)

COMMENT

Contact: Cushman JC
 Department of Biochemistry and Molecular Biology
 Oklahoma State University
 350 Noble Research Center, Stillwater, OK 74078-3035, USA
 Tel: 405-744-6207
 Fax: 405-744-7799
 Email: jcushman@biochem.okstate.edu

PCR Primers

FORWARD: T7
 BACKWARD: T3
 Plate: L0-4 row: G column: 2

Seq primer: M13 reverse
 High quality sequence stop: 300.

FEATURES

1. 402
 Location/Qualifiers
 /organism="Mesembryanthemum crystallinum"


```

Cp      22 CRTTNCRTTYTGRTGNARNGT 1

RESULT 3
LOCUS   AA962918          471 bp      mRNA           EST       18-MAY-1998
DEFINITION L30-449T3 Ice plant Lambda Uni-Zap XR expression library, 30 hour NaCl treatment Mesembryanthemum crystallinum cDNA clone L30-449 5' similar to 40S ribosomal protein S4 type I, mRNA sequence.
ACCESSION AA962918
NID       93136391
KEYWORDS  EST.
SOURCE    Common ice plant.
ORGANISM  Mesembryanthemum crystallinum
          Eukaryota; Viridiplantae; Charophyta/Embryophyta group; Embryophyta; Tracheophyta; seed plants; Magnoliophyta; eudicotyledons; Caryophyllidae; Caryophyllales; Alstroceae; Mesembryanthemum.
REFERENCE 1 (bases 1 to 471)
AUTHORS   Cushman, J.C.
TITLE      An expressed sequence tag database for the common ice plant, Mesembryanthemum crystallinum
JOURNAL   Unpublished (1997)
COMMENT   Contact: Cushman JC
          Department of Biochemistry and Molecular Biology
          Oklahoma State University
          350 Noble Research Center, Stillwater, OK 74078-3035, USA
          Tel.: 405-744-6207
          Fax: 405-744-7799
          Email: jcushman@biochem.okstate.edu
          PCR Primers
          FORWARD: T7
          BACKWARD: T3
          Plate: L30-5 row: E column: 5
          Seq primer: T3
          High quality sequence stop: 300.
          Location/Qualifiers
            1..471
              /organism="Mesembryanthemum crystallinum"
              /note="Vector: Lambda Uni-Zap XR, Bluescript SK-, Site_1:"
              ECOR1_Site_2: XhoI"
              /db_xref="taxon:3544"
              /clone="L30-449"
              /clone_lib="Ice plant Lambda Uni-Zap XR expression library, 30 hours NaCl treatment"
              /tissue_type="leaf, 30 h 0.4M NaCl"
              /dev_stage="Six week Old"
              /seq_id=111 c 101 g 134 t

BASE COUNT 125 a 111 c 101 g 134 t
ORIGIN
Query Match 95.0%; Score 95; DB 14; Length 471;
Best Local Similarity 63.6%; Pred. No. 3, 12e-04;
Matches 14; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Db 210 CATGGCATCTGATGCACAGT 231
|:||||||:::||:|:-|||
Cp 22 CRTTNCRTTYTGRTGNARNGT 1

RESULT 4
LOCUS   AA887369          472 bp      mRNA           EST       30-MAR-1998
DEFINITION L30-342T3 Ice plant Lambda uni-zap XR expression library, 30 hours NaCl treatment Mesembryanthemum crystallinum cDNA clone L30-342 5' similar to 40S ribosomal protein S4, type I, mRNA sequence.
ACCESSION AA887369
NID       93002461
KEYWORDS  EST.
SOURCE    Common ice plant.
ORGANISM  Mesembryanthemum crystallinum
          Eukaryota; Viridiplantae; Charophyta/Embryophyta group; Embryophyta; Tracheophyta; seed plants; Magnoliophyta; eudicotyledons; Caryophyllidae; Caryophyllales; Alstroceae; Mesembryanthemum.
```


FEATURES	SOURCE
Location/Qualifiers	1..486
/organism="Mesembryanthemum crystallinum"	
/note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site-1	
ECORI; Site-2: XhoI"	
/db_xref="taxon:3544"	
/clone="130-430"	
/clone_1lb="Ice plant Lambda Uni-Zap XR expression	
library, 30 hours NaCl treatment"	
/tissue_type="Leaf, 30 h 0.4M NaCl"	
/dev_stage="Six week old"	
129 a 115 c 107 g 134 t 1 others	

Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria.

Query Match	85.0%	Score 85;	DB 19;	Length 290;
Best Local Similarity	59.1%	Pred. No. 9.90e-02;		
Matches	13;	Conservative 5;	Mismatches 4;	Indels 0;
			Gaps	0

```

Db      78 CCTTCACCAGCATGGCAACGT 99
      | :| | | :| | :| | | |
QY      2 CNYTNCAYCARAAYGNAAYGT 23

```

RESULT	12			
LOCUS	T19059	293 bp	mRNA	
DEFINITION	b06005s Testis 2 Homo sapiens cDNA clone b06005 5' end, mRNA sequence.			25-SEP-1996
ACCESSION	U00600			

KEYWORDS	EST.
SOURCE	human.

ORGANISM Homo sapiens
Eukaryotae; Mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 293)
REFERENCE Pawlak, A., Toussaint, C., Levy, I., Bulle, F., Poyard, M., Barouki, R.
and Guellaen, G.

TITLE	Characterization of a large population of mRNAs from human testis
JOURNAL	Genomics 26, 151-158 (1995)
MEDLINE	95301283
COMMENT	

Contact: Guellaen G
Unité INSERM 99
INSERM

FEATURES	Location/Qualifiers
source	1. .293

/Organism:"Homo sapiens"/
"note:-Vector: pGEM 5zf(+); site_1: SalI; site_2: NotI;
cDNA was prepared from human testis of a 27 years old man
DNA was prepared using a 15mer oligo dT anchored by two
degenerated bases at its 3' end and containing a NotI site
at its 5' end. The cDNA was cloned between SalI and NotI
sites of pGEM 5zf(+). The 3' and 5' ends at the NotI site. The
SalI site is lost during the cloning procedure. cDNA
corresponding to abundant species were eliminated from
this library."

miRNA	<1.	.>293			
BASE COUNT	63 a	101 c	85 g	40 t	4 others
ORIGIN					

Query Match	85.0%;	Score 85;	DB 21;	Length 293;
Best Local Similarity	59.1%;	Pred. No. 9.90e-02;		
Matches	13;	Conservative	5;	Mismatches 4;
			Indels	0;
			Gaps	0

Db 105 CCTTCCACCAGCATGGCAACGT 12
| : | | : | : | : | | : | |
QY 2 CNYTNCAYCARAAYGGNAAYGT 23

RESULT	13			
LOCUS	T34361	305 bp	mRNA	EST
DEFINITION	EST666604 Homo sapiens CDNA 5' end similar to None.			06-SEP-1998
ACCESSION	T34361			
NID	9616459			
KEYWORDS	EST.			
SOURCE	human primer-M13 Reverse library-Human Prostate gland.			
ORGANISM	Homo sapiens			

REFERENCE AUTHORS

Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Hominoidea; Homo. 1 (bases 1 to 305)

TITLE	Initial Assessment of Human Gene Diversity and Expression Patterns
JOURNAL	Based Upon 52 Million Basepairs of cDNA Sequence
COMMENT	Unpublished (1995)
Other_ESTS:	THC9497

JOURNAL
COMMENT

The Institute for Genomic Research
932 Clopper Rd, Gaithersburg, MD 20878
Tel: 301.507.3000

Tel: 3018699056
Fax: 3018699423

For clone availability, additional sequence and expression information related to this EST, please contact the TIGR Database (tdbinfo@tigr.org).

FEATURES	Location/Qualifiers
source	1. .305

ORIGIN	BASE COUNT	mRNA	/organism="Homo sapiens"
75 a	84 c	92 g	52 t
2 others			

Query Match	85.0%;	Score 85;	DB 16;	Length 305;
Best Local Similarity	59.1%;	Pred. No. 9.90e-02;		
Matches	13;	Mismatches	4;	Indels 0;
				Gaps 0

```

Db      29 CCTTCACCAAGCATGGCAACGT 50
      | : | | | : | : | | | : |
QY      2 CNYTNCAYCARAAYGGNAAYGT 23

```

RESULT	14	D4E601	371	---
TOTALS		SUM		---

sapiens genomic clone Plate=CT 784 Col=12 Row=A, genomic survey sequence.

ACCESSION	B45681
NID	92550515

ORGANISM *Homo sapiens*

Eukaryotes; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria
Primates; Catarrhini; Homnidae; Homo.

REFERENCE

Authors	Title
Maniatis, G., Zaccarone, R.D., Smith, T., Tipton, S., Schmidt, S., Traicoff, R., Abajian, C., Blanchard, A., West, A. and Hood, L.E.	Construction of a Characterized Clone Resource for Genomic

JOURNAL
OF
MOLECULAR
EVOLUTION

JOURNAL
COMMENT

Contact: Mahairas GG, Zackrone KD, Hood L
University of Washington

Job time : 106 secs.

Seattle, WA 98195, USA
Tel: (206) 616-8744
Fax: (206) 685-7301

Email: kzackron@u.washington.edu
Sequence Tagged Connector

Plate: C1 784 row: A column: 12
Class: BAC ends

High quality sequence stop: 371.

FEATURES

source

1. 371

/organism="Homo sapiens"

/note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in E-Coli DH108"

/db_xref="taxon:9606"

/clone_plate="C1 784 Col-12 Row-A"

/clone_lib="C1 Human Genomic Sperm Library C"

/sex="M"

BASE COUNT

97 a 91 c 72 g 111 t

ORIGIN

Query Match

85.0%; Score 85; DB 26; Length 371;

Best Local Similarity 63.2%; Pred. No. 9.90e-02;

Matches 12; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

DB 286 CATGCCATTGATGTAG 304

CP 22 CTTNCCRTTGTGTGNAR 4

RESULT 15

LOCUS C83092 377 bp mRNA EST 17-JUN-1998

DEFINITION Oryctolagus cuniculus corneal endothelial CDNA, clone 1341, mRNA

sequence.

ACCESSION C83092

93062049

KEYWORDS EST: cornea; endothelium; rabbit.

SOURCE Oryctolagus cuniculus corneal endothelial cell CDNA to mRNA,

clone:1341.

ORGANISM Oryctolagus cuniculus

Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;

Lagomorpha; Leporidae; Oryctolagus.

REFERENCE Fujimaki,T., Hotta,Y., Sakuma,H., Fujiki,K. and Kanai,A.

AUTHORS Large Scale Sequencing of the Rabbit Corneal Endothelial CDNA

TITLE Library

JOURNAL Unpublished (1998)

REFERENCE 2 (bases 1 to 377)

AUTHORS Fujimaki,T.

JOURNAL Direct Submission

Submitted (01-DEC-1997) to the DDBJ/EMBL/GenBank databases. Takuro

Fujimaki, Juntendo University School of Medicine, Department of

Ophthalmology; 3-1-3, Hongo, Bunkyo-ku, Tokyo, Japan, Tokyo 113,

Japan (E-mail:fujimaki@med.juntendo.ac.jp, Tel:+81-3-5802-1092,

Fax:+81-3-3817-0260)

FEATURES 3 Location/Qualifiers

source 1. 377

/organism="Oryctolagus cuniculus"

/db_xref="taxon:9986"

/cell_type="corneal endothelial cell"

/clone="1341"

BASE COUNT 108 a 75 c 100 g 92 t 2 others

ORIGIN

Query Match 85.0%; Score 85; DB 15; Length 371;

Best Local Similarity 65.0%; Pred. No. 9.90e-02;

Matches 13; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

DB 182 ACATTCATCAGATGGCA 201

OY 1 ACNTNCAYCARAYGNAA 20

Search completed: Tue Apr 27 10:34:52 1999

2/

THIS PAGE BLANK (USPTO)

 M P E R E I E
 (TM)

Release 3.1a John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MParch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Fri Apr 16 14:18:00 1999; MasPar time 3.81 Seconds
 167.077 Million cell updates/sec
 Tabular output not generated.

Title: >US-09-011-797-2
 Description: (1-17) from US09011797.pep
 Perfect Score: 116
 Sequence: 1 FGFTGARKSARKLANQ 17

Scoring table:
 PAM 150
 Gap 15

Searched: 116695 segs, 37453910 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: p1r58
 1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 26.726; Variance 36.801; scale 0.726

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	116	100.0	109	2	S70631	5.09e-13
2	116	100.0	168	2	S60337	5.09e-13
3	116	100.0	176	2	JC6152	5.09e-13
4	116	100.0	181	2	JC6151	5.09e-13
5	116	100.0	187	2	JC4502	5.09e-13
6	67	57.8	205	2	S73721	7.35e-02
7	66	56.9	205	2	D64234	1.16e-01
8	61	52.6	405	1	DEPMNA	1.08e+00
9	61	52.6	405	1	DEPMNA	1.08e+00
10	60	51.7	310	2	A55768	1.66e+00
11	59	50.9	177	2	C69476	2.55e+00
12	59	50.9	211	2	A70307	2.55e+00
13	59	50.9	231	2	C71532	2.55e+00
14	58	50.0	208	2	F64681	3.89e+00
15	58	50.0	421	2	C69665	3.89e+00
16	58	50.0	998	2	A36368	3.89e+00
17	57	49.1	159	2	H65167	5.91e+00
18	57	49.1	528	2	B64760	5.91e+00
19	57	49.1	801	2	PC6010	5.91e+00
20	56	48.3	306	2	S39151	8.93e+01
21	55	47.4	173	2	S42246	1.34e+01
22	55	47.4	264	2	B64714	1.34e+01
23	55	47.4	470	2	S68964	1.34e+01

24	55	47.4	476	2	S68965	1.34e+01
25	54	46.6	224	2	F69091	2.01e+01
26	54	46.6	374	2	S41758	2.01e+01
27	54	46.6	475	2	S73746	2.01e+01
28	53	46.6	651	2	D70789	2.01e+01
29	53	45.7	93	2	S25976	2.99e+01
30	53	45.7	141	2	S73848	2.99e+01
31	53	45.7	300	2	A42204	2.99e+01
32	53	45.7	323	2	S42426	2.99e+01
33	53	45.7	345	2	E70794	2.99e+01
34	53	45.7	352	2	A69103	2.99e+01
35	53	45.7	378	1	A49662	2.99e+01
36	53	45.7	400	2	A71490	2.99e+01
37	53	45.7	475	1	WMA06	2.99e+01
38	53	45.7	482	2	S31478	2.99e+01
39	53	45.7	547	2	A64878	2.99e+01
40	53	45.7	692	2	H69691	2.99e+01
41	53	45.7	704	2	S46000	2.99e+01
42	53	45.7	856	2	A64699	2.99e+01
43	52	44.8	90	1	DNZRM	4.41e+01
44	52	44.8	125	2	S30170	4.41e+01
45	52	44.8	305	2	A44362	4.41e+01

ALIGNMENTS

RESULT	ENTRY	ENTRY TITLE	ORGANISM	DATE	ACCESSIONS	REFERENCE	#authors	#journal	#title	#accession	#status	#molecule-type	#residues	#cross-references	#length	#checksum
1	S70631	orphanin FQ/nociceptin - mouse (fragment)	mouse	14-Feb-1997	S70631	Pan, Y.X.; Xu, J.; Pasternak, G.W.	1		Cloning and expression of a cDNA encoding a mouse brain orphanin FQ/nociceptin precursor.	S70631	preliminary	molecule-type mRNA	1-109	EMBL:U44027; NID:q1335869; P1D:q1335870	109	9509
2	S60337	Opioide receptor-like ORL(1) receptor antagonist - rat (fragment)	Rattus norvegicus	06-Sep-1996	S60337	Meunier, J.C.; Mollereau, C.; Toll, L.; Snaudeau, C.; Moisset, C.; Alviere, P.; Butour, J.L.; Guillemot, J.C.; Ferrera, P.; Monsarrat, B.; Mazarguil, H.; Vassart, G.; Parmentier, M.; Costentin, J.	1		Isolation and structure of the endogenous agonist of opioide receptor-like ORL(1) receptor.	S60337	preliminary; not compared with conceptual translation	molecule-type mRNA	1-109	EMBL:U44027; NID:q1335869; P1D:q1335870	109	9509

##residues 1-168 ##label MEU
SUMMARY #length 168 #checksum 3964

Query Match 100.0%; Score 116; DB 2; Length 168;
Best Local Similarity 100.0%; Pred. No. 5.09e-13;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 122 FGFTGARKSARKLANQ 138
|||||
QY 1 FGFTGARKSARKLANQ 17

RESULT 3
ENTRY JC6152 #type complete
TITLE orphanin FQ precursor - human
ALTERNATE_NAMES nociceptin precursor
ORGANISM #formal_name Homo sapiens #common_name man
DATE 11-Apr-1997 #sequence_revision 09-May-1997 #text_change 10-Sep-1997
ACCESSIONS JC6152
REFERENCE JC6151
#authors Nothacker, H.P.; Reinscheid, R.K.; Mansour, A.; Henningsen, R.A.; Ardati, A.; Monsma Jr., F.J.; Watson, S.J.; Civelli, O.
#journal Proc. Natl. Acad. Sci. U.S.A. (1996) 93:8677-8682
#title Primary structure and tissue distribution of the orphanin FQ precursor.
#accession JC6152
##molecule_type mRNA
##residues 1-176 ##label NOT
#cross-references GB:U48263; NID:g1185009; PID:g1185010
COMMENT This protein is a neuropeptide that is an endogenous ligand to a G-protein-coupled receptor sequentially related to the opiod receptors. It functions in development.
KEYWORDS opiod peptide
SUMMARY #length 176 #molecular-weight 20295 #checksum 498

Query Match 100.0%; Score 116; DB 2; Length 176;
Best Local Similarity 100.0%; Pred. No. 5.09e-13;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 130 FGFTGARKSARKLANQ 146
|||||
QY 1 FGFTGARKSARKLANQ 17

RESULT 4
ENTRY JC6151 #type complete
TITLE orphanin FQ precursor - rat
ALTERNATE_NAMES nociceptin precursor
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 11-Apr-1997 #sequence_revision 09-May-1997 #text_change 10-Sep-1997
ACCESSIONS JC6151
REFERENCE JC6151
#authors Nothacker, H.P.; Reinscheid, R.K.; Mansour, A.; Henningsen, R.A.; Ardati, A.; Monsma Jr., F.J.; Watson, S.J.; Civelli, O.
#journal Proc. Natl. Acad. Sci. U.S.A. (1996) 93:8677-8682
#title Primary structure and tissue distribution of the orphanin FQ precursor.
#accession JC6151
##molecule_type mRNA
##residues 1-181 ##label NOT
#cross-references GB:U48262; NID:g1185011; PID:g1185012
COMMENT This protein is a neuropeptide that is an endogenous ligand to a G-protein-coupled receptor sequentially related to the opiod receptors. It is involved in development.
KEYWORDS opiod peptide
SUMMARY #length 181 #molecular-weight 20478 #checksum 3303

Query Match 100.0%; Score 116; DB 2; Length 181;
Best Local Similarity 100.0%; Pred. No. 5.09e-13;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 135 FGFTGARKSARKLANQ 151
|||||
QY 1 FGFTGARKSARKLANQ 17

RESULT 5
ENTRY JC4502 #type complete
TITLE opiod-receptor like G protein coupled receptor
ALTERNATE_NAMES nociceptin/orphanin FQ protein precursor - mouse
ORGANISM N23K protein; nociceptin precursor homolog
#formal_name Mus musculus #common_name house mouse
DATE 07-Feb-1996 #sequence_revision 19-Apr-1996 #text_change 10-Sep-1997
ACCESSIONS JC4502
REFERENCE JC4502
#authors Saito, Y.; Maruyama, K.; Saito, T.C.; Kawashima, S.
#journal Biochem. Biophys. Res. Commun. (1995) 217:539-545
#title N23K, a gene transiently up-regulated during neural differentiation, encodes a precursor protein for a newly identified neuropeptide nociceptin.
#accession JC4502
##molecule_type mRNA
##residues 1-187 ##label SAI
#cross-references DDBJ:DS0056; NID:g1857023; PID:di009395; PID:g1217906
REFERENCE JC4502
#authors Houtani, T.; Nishi, M.; Takeshima, H.; Nukada, T.; Sugimoto, T.
#journal Biochem. Biophys. Res. Commun. (1996) 219:714-719
#title Structure and regional distribution of nociceptin/orphanin FQ precursor.
#accession JC4502
##molecule_type mRNA
##residues 1-187 ##label HOU
#cross-references DDBJ:DS2866; NID:g1311472; PID:di012281; PID:g1311473
#experimental_source brain
#note The authors translated the codon CTC for residue 72 as Lys and GAC for residue 76 as Glu
COMMENT This protein belongs to the opiate precursor family, and functions as a neuropeptide precursor and plays a role in neuronal differentiation and development.
GENETICS N23K
#gene brain; differentiation; G protein-coupled receptor; neuropeptide
FEATURE 1-187
19-187
109-126 #domain signal sequence #status predicted #label SIG
141-157 #product neuropeptide precursor, N23K #status predicted
160-176 #label MAP
#region D-A-E-P-G-A motif
#product neuropeptide #status predicted #label NRP
#product heptadecapeptide #status predicted #label HPP
SUMMARY #length 187 #molecular-weight 20884 #checksum 8758

Query Match 100.0%; Score 116; DB 2; Length 187;
Best Local Similarity 100.0%; Pred. No. 5.09e-13;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 141 FGFTGARKSARKLANQ 157
|||||
QY 1 FGFTGARKSARKLANQ 17

RESULT 6
ENTRY S73721 #type complete
TITLE ribosomal protein S4 - Mycoplasma pneumoniae (ATCC 29342)
ALTERNATE_NAMES (SGC3)
ORGANISM hypothetical protein H08_orf205
#formal_name Mycoplasma pneumoniae
#variety ATCC 29342
DATE 27-Feb-1997 #sequence_revision 25-Apr-1997 #text_change

21-Aug-1998
ACCESSIONS S73721
REFERENCE S73327
#authors Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkl, E.; Li,
B.C.; Herrmann, R.
#journal Nucleic Acids Res. (1996) 24:4420-4449
#title Complete sequence analysis of the genome of the bacterium
Mycoplasma pneumoniae.
#cross-references MIMD:97105885
#accession S73721
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-205 ##label HIM
##cross-references EMBL:AE000038; GB:U00089; NID:g1674074; PID:g1674077
#note the nucleotide sequence was submitted to the EMBL Data
Library, November 1996

GENETICS
#gene rpsD
#genetic_code SGC3
CLASSIFICATION #superfamily Escherichia coli ribosomal protein S4
KEYWORDS protein biosynthesis; ribosome
SUMMARY #length 205 #molecular-weight 23774 #checksum 29

Query Match 57.8%; Score 67; DB 2; Length 205;
Best Local Similarity 53.3%; Pred. No. 7.35e-02;
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 105 GFAPTRRSARQVNH 119
QY 3 GFTGARSARKLANQ 17
||:|||||:|
|:|:|:|:|:|

RESULT 7
ENTRY D64234 #type complete
TITLE ribosomal protein S4 - Mycoplasma genitalium (SGC3)
#formal_name Mycoplasma genitalium
DATE 17-Nov-1995 #sequence_revision 17-Nov-1995 #text_change
21-Aug-1998

ACCESSIONS D64234
ENTRY A64200
#authors Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton,
R.A.; Fleischmann, R.D.; Bult, C.J.; Kerlavage, A.R.;
Sutton, G.; Kelley, J.M.; Fritchman, J.L.; Weidman, J.F.;
Small, K.V.; Sandusky, M.; Fuhmann, J.; Nguyen, D.;
Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick,
J.M.; Tomb, J.F.; Dougherty, B.A.; Bort, K.F.; Hu, P.C.;
Lucier, T.S.; Peterson, S.N.; Smith, H.O.; Hutchison III,
C.A.; Venter, J.C.
#journal Science (1995) 270:397-403
#title The minimal gene complement of Mycoplasma genitalium.
#cross-references MIMD:96026346
#accession D64234
#status preliminary; nucleic acid sequence not shown;
translation not shown

##molecule_type DNA
##residues 1-205 ##label TIGR
##cross-references GB:U39712; GB:U43967; NID:g1046005; PID:g1046011;
TIGR:M311

##experimental_source strain G-37

GENETICS
#genetic_code SGC3
CLASSIFICATION #superfamily Escherichia coli ribosomal protein S4
SUMMARY #length 205 #molecular-weight 23947 #checksum 310

Query Match 56.9%; Score 66; DB 2; Length 205;
Best Local Similarity 53.3%; Pred. No. 1.16e-01;
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 105 GFAPTRRSARQVNH 119
QY 3 GFTGARSARKLANQ 17
||:|||||:|
|:|:|:|:|:|

RESULT 8
ENTRY S74025 #type complete
TITLE conserved hypothetical protein c0207 - Sulfolobus
solifarius
ORGANISM Sulfolobus solifarius
#formal_name Sulfolobus solifarius
DATE 09-Oct-1997 #sequence_revision 24-Oct-1997 #text_change
21-Aug-1998

ACCESSIONS S74025
ENTRY S73076
#authors Sensen, C.W.; Klenk, H.P.; Singh, R.K.; Allard, G.; Chan,
C.C.Y.; Liu, Q.Y.; Penny, S.L.; Young, F.; Schenk, M.E.;
Gasterland, T.; Doolittle, W.F.; Ragan, M.A.; Charlebois,
R.L.
#journal Mol. Microbiol. (1996) 22:175-191
#title Organizational characteristics and information content of an
archaeal genome: 156 kb of sequence from Sulfolobus
solifarius p2.
#accession S74025
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-526 ##label SEN
##cross-references EMBL:Y08256; NID:g1707679; PID:e283902; PID:g1707719
#experimental_source strain p2
#note the nucleotide sequence was submitted to the EMBL Data
Library, September 1996

CLASSIFICATION #superfamily proline carrier protein
SUMMARY #length 526 #molecular-weight 56210 #checksum 8544

Query Match 56.9%; Score 66; DB 2; Length 526;
Best Local Similarity 52.9%; Pred. No. 1.16e-01;
Matches 9; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Db 220 FGFGSAGHNAQTLSQ 236
QY 1 FGFTGARSARKLANQ 17
|||||:|:|:|
|:|:|:|:|:|

RESULT 9
ENTRY DEPMNA #type complete
TITLE glyceraldehyde-3-phosphate dehydrogenase (NADP+)
(phosphorylating) (EC 1.2.1.13) A precursor, chloroplast -
garden pea
ORGANISM Pisum sativum #common_name garden pea
DATE 05-Sep-1997
#formal_name Pisum sativum #common_name garden pea
31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change
21-Aug-1998

ACCESSIONS S14243; S05554; A38270
ENTRY A38270
#authors Lland, M.F.; Zhang, D.X.; Cerff, R.
#journal Proc. Natl. Acad. Sci. U.S.A. (1990) 87:8918-8922
#title Differential intron loss and endosymbiotic transfer of
chloroplast glyceraldehyde-3-phosphate dehydrogenase genes
to the nucleus.
#cross-references MIMD:91062390
#accession S14243
#molecule_type DNA
#residues 1-405 ##label LIA
##cross-references EMBL:X52148; NID:g12158; PID:g12159
S05552

REFERENCE S05552
#authors Brinkmann, H.; Cerff, R.; Salomon, M.; Soll, J.
#journal Plant Mol. Biol. (1989) 13:81-94
#title Cloning and sequence analysis of cDNAs encoding the cytosolic
precursors of subunits gapa and gapb of chloroplast
glyceraldehyde-3-phosphate dehydrogenase from pea and
spinach.
#accession S05554
#molecule_type mRNA
#residues 1-405 ##label BRI
##cross-references EMBL:X15190
#note the authors translated the codon AGA for residue 194 as
Gly; the sequence shown follows the authors' translation

GENETICS
#gene Gpal


```
#common trans\ peptide (chloroplast) #status predicted
#label
#product glyceraldehyde-3-phosphate dehydrogenase
#(NDP+) (phosphorylating) A #status predicted #label
MAT\
```

```
#active_site cys, His #status predicted
#length 405 #molecular-weight 43338 #checksum 1819
```

Query Match	52.68;	Score 61;	DB 1;	Length 405,
Best Local Similarity	46.78;	Pred. No. 1.08e+00;		
Matches	7;	Conservative	5;	Mismatches 3;
				Indels

D_b 20 FSEFSGLRNSSRHL_P 34

QY 1 EGGFTGARKSARKLA 15

ENTRY	TITLE	ALTERNATE_NAMES	ORANISM	DATE
A55768	#type complete	asparaginyl-peptide amidohydrolase (EC 3.5.1.-) - pig asparagine amidohydrolase		
		asparagine amidohydrolase		
		#Formal_name Sus scrofa domestica #common_name domestic pig		
		2-Mar-1995 #sequence revision 23-Mar-1995 #text change		

#authors Stewart, A.E.; Arfin, S.M.; Bradshaw, R.A.
#journal J. Biol. Chem. (1995) 270:25-28
#title The sequence of porcine protein NH-2-terminal asparagine

```
#accession      A55768
#status         preliminary
#molecule_type mRNA
#residues       1-310 #label STE
#cross-references GB:U1062; NID:g555949; PID:g555950
KEYWORDS
hydrolyase
```

```
#product asparaginyl-peptidase amidohydrolase #status
      predicted #label MAR
#length 310 #molecular-weight 34760 #checksum 7268
```

Query Match	51.78;	Score 60;	DB 2;	Length 310;
Best Local Similarity	43.88;	Pred. NO. 1.66e+00;		
Matches	7;	Conservative	5;	Mismatches 4;
			Indels	0;
			Gaps	0;

|||: | ::||::|
2 GGFTGARKSARKLANQ 17

ENTRY	#type complete
TITLE	hypothetical protein AF1812 - Archaeoglobus fulgidus
ORGANISM	#normal_name Archaeoglobus fulgidus
DATE	05-Dec-1997 #sequence_revision 05-Dec-1997 #text_cha
	05-Jun-1998

K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.; Peterson, J.D.; Richardson, D.L.; Kerlavage, A.R.; Graham.

Dougherty, B.A.; Mckenney, K.; Adams, M.D.; Loftus, B.; Peterson, S.; Reich, C.I.; McNeill, L.K.; Badger, J.H.; Gloder, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.; Uterback, T.; Cotton, M.D.; Spilligs, T.; Attisch, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.; D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.;

#journal Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese
 #title C.R.; Venter, J.C.
 Nature (1997) 390:364-370
 The complete genome sequence of the hyperthermophilic,
 sulfate-reducing archaeon *Archaeoglobus fulgidus*.
 #cross-references WUID:98049343
 #accession 069476

```

#status      preliminary; nucleic acid sequence not shown;
              translation not shown

```

```
#cross-references GB:AE000978; GB:AE000782; NID:g2689301; PID:g2648748;
TIGR:AF1812
```

Query Match	50.9%;	Score 59;	DB 2;	Length 177;
Best Local Similarity	46.7%;	pred NO	2.55e+00.	

conservative; models; gaps

1 FGGFTGARKSARKLA 15

```

ENTRY
TITLE
ABSTRACT
KEYWORDS
ORGANISM
DATE
CROSSREFS
A70307 #type complete
        ribosomal protein S04 - Aquifex aeolicus
#formal_name Aquifex aeolicus
08-May-1998 #sequence_revision 08-May-1998 #text_change
02-Jul-1998
170307

```

#authors	#journal	#title
Dekker, G.; Warren, P. V.; Gaasterland, T.; Young, W. G.; Lenox, A. L.; Graham, D. E.; Overbeek, R.; Shead, M. A.; Keller, M.; Anjaly, M.; Huber, R.; Feldman, R. A.; Short, J. M.; Olson, G. J.; Swanson, R. V.	Nature	The complete genome of the hyperthermophilic bacterium <i>Aquifex aeolicus</i> .

```
##status      preliminary; nucleic acid sequence not shown;
               translation not shown
```

```

#cross_references GB:AE000672; NID:g2982810; PID:g2982819; GB:AE000657
#experimental_source strain VFS
GENETICS

```

CLASSIFICATION	
#superfamily	Escherichia coli ribosomal protein S4
#length	211
#molecular-weight	24815
#checksum	7435

Query Match 50.9%; Score 59; DB 2; Length 211;
Best Local Similarity 40.0%; Pred. No. 2.55e+00;
Matches 6; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

```

      3 GETGARKSARKLANQ 17
      ||:::|: || |

```

ENTRY TITLE	#type complete probable ribonuclease III - Chlamydia trachomatis (serotype D, strain UW3/Cx)
ORGANISM SITE	#formal_name Chlamydia trachomatis 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change

ACCESSIONS	
REFERENCE	
#authors	
C71532	Stephens, R.S.; Kaiman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, W.P.; Olinger, L.; Taturov, R.L.; Zhao, Q.; Koonin, E.V.; Davis, R.W.
A71460	


```

#submission      Submitted to GenBank      May 1998
#description      Genome sequence of an obligate intracellular pathogen of
                  humans: Chlamydia trachomatis.

#accession      C71532
#status          Preliminary
##molecule_type DNA
##residues       1-231 ##label ARN
##cross-references GB:AE001302; GB:AE001273; NID:g33z8708; PID:g33z8712
##experimental_source serotype D, strain UW-3/Cx

GENETICS
#gene            inc
SUMMARY          #length 231 #molecular-weight 25561 #checksum 9834

Query Match      50.9%; Score 59; DB 2; Length 231;
Best Local Similarity 61.5%; Pred. No. 2,55e+00;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 207 GFAGSKKEAEKLA 219
|||:|:| | | | |
3 GFTGARKSARKLA 15

RESULT 14
ENTRY           F64681 #type complete
TITLE           ribosomal protein S4 - Helicobacter pylori (strain 26695)
ORGANISM        #format_name Helicobacter pylori
DATE            09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change
                31-Oct-1997
ACCESSIONS      F64681
REFERENCE        AG4520
#authors        Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.;
                Sutton, G.G.; Fleischmann, R.D.; Ketchum, K.A.; Klenk,
                H.P.; Gill, S.; Dougherty, B.A.; Nelson, K.; Quackenbush,
                J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus, B.;
                Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.;
                McKenney, K.; Fitzgerald, L.M.; Lee, N.; Adams, M.D.;
                Hickey, E.K.; Berg, D.E.; Gocayne, J.D.; Utterback, T.R.;
                Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.;
                Fujii, C.; Bowman, C.; Watthey, L.; Wallin, E.; Hayes,
                W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,
                C.M.; Venter, J.C.
                Nature (1997) 388:539-547
                The complete genome sequence of the gastric pathogen
                Helicobacter pylori.
#cross-references M01D:97394467
#accession      F64681
#status          preliminary; nucleic acid sequence not shown;
                translation not shown

##molecule_type DNA
##residues       1-208 ##label TOM
##cross-references GB:AE000633; GB:AE000511; NID:g2314452; PID:g2314460.
                TIGR:HP1294
CLASSIFICATION  #superfamily Escherichia coli ribosomal protein S4
SUMMARY          #length 208 #molecular-weight 23964 #checksum 6692

Query Match      50.0%; Score 58; DB 2; Length 208;
Best Local Similarity 46.7%; Pred. No. 3,89e+00;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Db 108 GFATRSSARQLVTH 122
|||:| | | | | :
3 GFTGARKSARKLANQ 17

RESULT 15
ENTRY           C69665 #type complete
TITLE           nitrate transporter nasa - Bacillus subtilis
ORGANISM        #format_name Bacillus subtilis
DATE            05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
                24-Sep-1998
ACCESSIONS      C69665; I40026
REFERENCE        AG9580
#authors        Kustu, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;

```

```

Alloni, G.; Acevedo, V.; Bertero, M.G.; Bessieres, P.;
Bolton, A.; Borchert, S.; Bollers, R.; Boursier, L.; Brans
A., Braun, M.; Brignell, S.C.; Bron, R.; Brouillet, S.;
Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;
Ehrlich, S.D.; Emerson, P.T.; Ertlan, K.D.; Errington, J.;
Fabret, C.; Ferrari, E.; Fouger, D.; Fitt, C.; Fujita,
M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Galleron, N.; Ghim,
S.Y.; Glaser, P.; Goffeau, A.G.; Goldtly, E.J.; Grand, G.;
Guisepi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood,
C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
Kasahara, Y.; Klauer-Blanchard, M.; Klein, C.; Kobayashi,
Y.; Koeter, P.; Konungstein, G.; Krogh, S.; Kumano, M.;
Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno
M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly
M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro,
V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott,
A.M.; Presecan, E.; Pulic, P.; Punnett, B.; Rapoport, G.;
Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Roche, E.;
Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon, E.;
Schleich, S.; Schroeter, R.; Scoffone, F.; Seikguchi, J.;
Sekorska, A.; Seror, S.J.; Serro, P.; Shin, B.S.; Soldo,
B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Tanahashi, H.;
Takekura, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
Vandanol, R.; Vannier, F.; Vassarotti, A.; Viari, A.;
Wambutt, R.; Wedler, E.; Wedler, H.; Weltzenegger, T.;
Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasumoto
K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumbstein, E.;
Yoshikawa, H.; Danchin, A.
Nature (1997) 390:249-256
The complete genome sequence of the Gram-positive bacterium
Bacillus subtilis.
#accession GB:299105; GB:AL009126; NID:g2632457; PID:e1182285;
#molecule_type DNA
#residues 1-421 #label KUN
#cross-references GB:299105; GB:AL009126; NID:g2632457; PID:g2632619
#experimental_source strain 168
REFERENCE
#authors Ogasawara, N.; Fujita, Y.; Kobayashi, Y.; Sadate, Y.; Tanaka
T.; Tanahashi, H.; Yamane, K.; Yoshikawa, H.
#journal Microbiology (1995) 141:257-259
#title Systematic sequencing of the Bacillus subtilis genome:
progress report of the Japanese group.
#cross-references MIMD:95219077
#accession I40026
#status translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 'L', 1-421 #label RES
#cross-references GB:D30689; NID:g710016; PID:g710017
GENERICs
#gene nash
#length 421 #molecular-weight 46067 #checksum 556
SUMMARY
Query Match 50.0%; Score 58; DB 2; Length 421;
Best Local Similarity 61.5%; Prod. No. 3.89e+00;
Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
Db 388 GKGAGKAAARM 400
II :|||||:
OY 2 GGTGAKSARKL 14

```


Wed Apr 28 09:08:33 1999

US-09-011-797-2.rpr

```

"search_tpn    n.a. - n.a. Smith-Waterman search, using a protein query

```

which has been backtranslated into n.a. using IUPAC symbols

Run on: Tue Apr 27 10:35:11 1999; masPar time 22.17 Seconds

Tabular output not generated.

Title: >US-09-011-797-4

Perfect Score: 100

Comp: TGNRANGTGGTYTTRCCNTTRCAN

Scoring table: TABLE bktranslated2

```
Nmatch      STD :  Dbase 0;  Query 0
```

Searched: 188442 seqs, 6802649 bases x 2

Post-processing: Minimum Match 0%

Database: n-geneseq32

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39 40:part40

Statistics: Mean 30.944; Variance 103.177; scale 0.300

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

Result	Query
--------	-------

No.	Score	Match Length	DB	ID	Description	Pred. No
1	100	100.0	932 34	T79888	DNA encoding human op	1.11e+00
2	85	85.0	2158 3	A14624	Plasmid pATPBD Inscr	2.77e+01
3	85	85.0	2158 21	T34372	Plasmid pATPBD (ATCC	2.77e+01
4	80	80.0	857	N93084	Lambda gt 11 p1(131)	7.71e+01
5	80	80.0	1763 3	A14071	kpnI-XbaI fragment fr	7.71e+01
6	80	80.0	1763 1	Q03242	Gene encodes function	7.71e+01
7	80	80.0	2839 1	Q03732	Human SKI related gen	7.71e+01
8	80	80.0	2888 1	Q03743	Human SKI related gen	7.71e+01
9	80	80.0	5687 29	T65501	Mouse cell cycle regu	7.71e+01
10	75	75.0	339 13	O68133	Sequence encoding hum	2.09e+02
11	75	75.0	339 13	O68132	Sequence encoding hum	2.09e+02
12	75	75.0	339 13	O68134	Sequence encoding hum	2.09e+02

C	13	75	75.0	339.13	068135	Sequence encoding hum	2.09e+020
C	14	75	75.0	339.13	068115	Sequence encoding hum	2.09e+020
C	15	75	75.0	339.13	068146	Sequence encoding hum	2.09e+020
C	16	75	75.0	1315.30	T66846	Soybean peroxidase SE	2.09e+020
C	17	75	75.0	1435.17	T66777	Bacillus sp. alkaline	2.09e+020
C	18	75	75.0	1462.10	066070	Sequence encoding soy	2.09e+020
C	19	75	75.0	2087.31	T69737	Corn starch branching	2.09e+020
C	20	75	75.0	2097.38	V01546	Rat syntaxin 1a gene	2.09e+020
C	21	75	75.0	2100.3	013985	Gamma-cyclodextrin gl	2.09e+020
C	22	75	75.0	2165.31	T69736	Corn starch branching	2.09e+020
C	23	75	75.0	2235.1	N90389	DNA encoding polypept	2.09e+020
C	24	75	75.0	2426.15	090271	Saccharomyces sp. rec	2.09e+020
C	25	75	75.0	2665.31	T69729	Plasmid pBE240 insert	2.09e+020
C	26	75	75.0	2855.39	V00014	Human sex comb on mid	2.09e+020
C	27	75	75.0	2855.39	V00011	Human sex comb on mid	2.09e+020
C	28	75	75.0	3065.39	V00014	Human sex comb on mid	2.09e+020
C	29	75	75.0	3065.39	V00010	Mouse sex comb on mid	2.09e+020
C	30	75	75.0	3097.31	T69590	Mouse sex comb on mid	2.09e+020
C	31	75	75.0	3255.39	T97299	Human sex comb on mid	2.09e+020
C	32	75	75.0	3255.39	V00013	Human sex comb on mid	2.09e+020
C	33	75	75.0	3327.39	V00012	Human sex comb on mid	2.09e+020
C	34	75	75.0	3327.39	T97298	Human sex comb on mid	2.09e+020
C	35	75	75.0	3337.1	N91089	Fragment of Pseudom	2.09e+020
C	36	75	75.0	4355.37	T89387	Bovine coronavirus E2	2.09e+020
C	37	75	75.0	4355.37	010947	Bovine coronavirus E2	2.09e+020
C	38	75	75.0	5255.3	025448	Marek disease virus g	2.09e+020
C	39	75	75.0	5547.35	T68844	Photobacterium luminesc	2.09e+020
C	40	75	75.0	6161.32	T77865	E.coli alkaline phosph	2.09e+020
C	41	75	75.0	7551.31	T66843	Photobacterium luminesc	2.09e+020
C	42	75	75.0	8585.28	T48667	Brassica napus micros	2.09e+020
C	43	75	75.0	8585.1	005749	Microspora genitalis c	2.09e+020
C	44	75	75.0	580073	T58840	Mycoplasma specific	2.09e+020
C	45	70.0	4287.40	V20475	Human AML1/Mt18 Oncog	5.47e+02	

ALIGNMENTS

RESULT	1	
ID	T79888	standard; DNA: 932 BP.
AC	T79888;	
DT	08-DEC-1997	(first entry)
DE	DNA encoding human opioid receptor-like 1 receptor ligand.	
KM	Opoid receptor-like 1; ORL1; ligand; pronociceptive; stress;	
KW	hyperalgesia; locomotor activity disease; anxiety; neuroendocrine;	
KW	memory; attention; sensory perception; learning; homeostasis;	
KW	hyperalgesia; nociception; endorphin; dynorphin A; ss.	
OS	Synthetic.	
FH	Key	Location/Qualifiers
FT	cds	1..507
FT		/*tag= a
FT		/note= "no start codon"
PN	W09707208-A1.	
PD	27-FEB-1997.	
PF	14-AUG-1996; BE0087.	
PR	15-AUG-1995; US-002368.	
PA	(ULBR) UNIV LIBRE BRUXELLES.	
PI	Neuner J, Mollereau C, Parmentier M, Vassart G;	
DR	MP1: 97-165292/15.	
DR	P-PSDB: W25162, W25163, W25164.	
PT	Novel ligand for the opiod receptor-like receptor, nociceptin -	
PT	has pro-nociceptive properties, useful for treating or preventing	
PT	diseases related to e.g. stress, hyperalgesia, locomotor activity,	
PT	etc.	
PS	Claim 1: Page -7: 48pp: English.	
CC	T79888 encodes a ligand to human opioid receptor-like 1 (ORL1) receptor	
CC	designated nociceptin, which resembles the endorphin dynorphin A.	
CC	Inhibitors of nociceptin or its DNA, e.g. antisense sequences and	
CC	antibodies can be used as new types of drugs in the control of various	
CC	behaviours or functions. The inhibitors can be used to prevent or treat	
CC	a disease related to hyperalgesia, neuroendocrine secretion, stress,	
CC	locomotor activity, anxiety, insinctive behaviour, learning and memory	
CC	homeostasis, hyperalgesia, hyperalgesia and/or sensory perception.	
Sequence	932 BP; 224 A; 273 C; 238 G; 197 T;	

Query Match	100.0%;	Score 100;	DB 34;	Length 932;
Best Local Similarity	65.2%;	Pred.	No. 1,1le+00;	
Matches 15;	Conservative	5;	Mismatches 3;	Indels 0; Gaps 0;
Db	481 actctgcaccagaatggtcaatgt 503			
	.:			
oy	1 ACNNTMCAYCCAAATGAGNAATGT 23			

RESULT	2
ID	Q14624 standard; CDNA; 2158 BP.
AC	Q14624;
DT	30-JAN-1992 (first entry)
DE	Plasmiid pPRAPDP insert encoding a dunce-like phosphodiesterase.
KW	rat brain; ATCC # 68586; cyclic nucleotide; PDE; ss.
Rattus	rattus.

Accession	Location/Qualifiers
U00001	1..1689
U00002	/*tag= a
U00003	W09116457-A.
U00004	31-OCT-1991.
U00005	19-APR-1991.
U00006	20-APR-1980; US-511715.
U00007	(COLD-) COLD SPRING HARBOR.
U00008	Wigler MH, Collicelli JJ;
U00009	WPL; 91-339841/46.
U00010	P-PSDB; R14836.
U00011	Complementary screening for genes and prod.s. - e.g. RAS protein
U00012	and cAMP, that modify complement or suppress genetic defect and
U00013	correct associated phenotypic alteration
U00014	Claim 15; Page 58; 16pp; English.
U00015	Plasmid pRADPD was isolated from a rat brain cDNA library. A large
U00016	open reading frame was found and a 562 amino acid sequence deduced
U00017	from it. The coding sequence is homologous to the Drosophila dunce
U00018	gene; the proteins encoded by both sequences have 80 per cent identic
U00019	without the introduction of gaps, over a 252 amino acid region.
U00020	(The dunce gene encodes a cyclic nucleotide PDE).
U00021	Sequence 2158 bp. 6233

Query Match	85.0%	Score 85;	DB 3;	Length 2158;
Best Local Similarity	59.1%;	Pred. No. 2.77e+01;		
Matches	13;	Conservative	5;	Mismatches 4;
			Indels	0;
			Gaps	0;

```

Db      670 cattaccattctgatgtgcgt 691
      1:| | | | | | | | : | |
Cp      22 CRTTNCRTTYTGRTGNARNGT 1

```

RESULT	3
ID	T34372 standard; CDNA; 2158 BP.
ACC	U34372.

09-OCT-1996 (first entry)
Plasmid PRATDPD (ATCC 68586) insert.
Bat brain for virus isolation

expression vector; pMDNS, RAS(vall9), strain TK161-R2V, *Drosophila*; heat shock sensitivity; PRATPD; alternate splicing; rat DPD cDNA; high affinity cAMP phosphodiesterase; ss.

Key	Location/Qualifiers
cds	1-2701

```

:1:2/02
/*tag= a
/product= Duncne-like phosphodiesterase
US5527896-A.

```

18-JUN-1996.
20-APR-1990;

20-APR-1990; US-511715.
19-APR-1991; US-688352.

MA {COLD-} COLD SPRING HARBOR LAB.
PI Colicelli JJ, Wigler MH;
BR WPT: 96-388803/30

P-PSDB; W00090.
DNA mols. isolated

glioblastoma cells - encode

PT RAS-related or cyclic nucleotide phosphodiesterase proteins
PS Claim 4; Column 29-34; 101BP; English.
CC This sequence represents a plasmid fragment which contains a rat brain
CC derived cDNA inserts encoding a dunce-like phosphodiesterase. The
CC cDNA brain cDNA library was cloned into the yeast expression vector
CC PAN3. DNA prepared from the cDNA expression library was used to
CC transform the RAS(yall9) yeast strain TK161-R2V. Only one transformant
CC displayed heat shock sensitivity which is conditional on retention of
CC from the transformant. The plasmid designated pRRDPD was isolated
CC from the transformant and sequenced. A large open reading frame of 562
CC codons was found. The first ATG appears at codon 46 and a protein which
CC initiates at this codon would have a predicted mol. wt. of approx. 60
CC kD. The isolated sequence was found to have homology to the *Drosophila*
CC dunce gene. The encoded proteins have an 808 amino acid identity
CC without the introduction of gaps over a 252 amino acid region located in
CC the centre of the rat DPD cDNA. The dunce gene has been shown to encode
CC a high affinity cAMP phosphodiesterase. The established open reading
CC frame in this sequence remains open for an additional 116 codons after
CC the termination codon, followed by more terminator codons, a
CC polyadenylation signal and a polyadenine stretch. This 3' open reading
CC frame could be incorporated into another dunce-like phosphodiesterase
CC through alternate splicing.
Sequence 2158 BP; 623 A; 536 C; 516 G; 483 T.

Query Match	85.0%	Score 85;	DB 21;	Length 2158;
Best Local Similarity	59.1%	Pred. No.	2.77e+01;	
Matches	13;	Conservative		

	Conservative	Mismatches	Indels	Gaps
Db 670	catcaatctcgtatgctg 691	1:1111111111111111		
Cp 22	CRITNCCRTTYTGRTGNARNGT 1			

RESULT	
ID	N93084 standard; CDNA; 857 BP.
NC	NC03084

AC	29-JUN-1990 (first entry)	CDNA encoding Der P I antigen
DT	Lambda gt 11 pl(13T) contg.	Der p I, house dust mite; allergy; asthma; ss.
DE		
KM		
NC	N53084;	

```

cds      Location/Qualifiers
key      Location/Qualifiers
FT       1..30
FT       /*tag= a
FT

```

```

FTT /product= prepro Der p I
FTT 31...735
FTT /*tag= b
FTT /product= pro Der p I
cdfs

```

```
cus      /2../35
FFFT     /*tag= c
FFFT     /product= mature Der p I
FFFT     803..808
FFFT     44..
```

WO8810297-A.
29-DEC-1988.
17-JUN-1988;
17-JUN-1987.

PRIN-) Princess Margaret Children's Med Res Found, (LUDW-) Ludwig Inst
Cancer Res, (HALL-) Walter and Eliza Hall Inst of Med Res,
I Thomas WR, Stewart GA, Turner KJ, Simpson RJ:
WDR-89-0029847002

Cloning of mite allergens from dermatophagoides - for obtaining polypeptide for use as diagnostic reagents and as therapeutic substances for allergic diseases

The cDNA encodes Der p 1, a 27 kD glycoprotein found in mite faeces which elicits IgE anti-mite antibody causing asthma, rhinitis, and atopic dermatitis. The class

subcloned into vector pGEX and expressed as a fusion protein with a glutathione transferase molecule. The polypeptide can be used as a diagnostic reagent since it reacts with IgE from serum of patients with allergic rhinitis.

It can also be used for desensitisation treatment.

CC See also N93085.
SQ Sequence 857 BP: 294 A: 169 C: 154 G: 240 T:

Query Match 80.0%; Score 80; DB 1; Length 857;
Best Local Similarity 57.1%; Pred. No. 7.71e+01;
Matches 12; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Db 135 cactccatctcgatgcaagg 155

1:|||||:|||||:|
Cp 22 CRTNCCRTTYTGRTGRRNG 2

RESULT 5

ID 014071 standard; DNA: 1763 BP.

AC 014071.
DT 10-JAN-1992 (first entry)
DE KpnI-XbaI fragment from pBV503 in Brevibacterium stationis IFO12144.
KM pCR30; pCR30; genetic engineering; ss.
OS Brevibacterium stationis IFO 12144.
PN J03210384-A.

PD 13-SEP-1991.
PF 11-JAN-1990: 004212.

PA (MIRP) MITSUBISHI PETROCH KK.
DR WPI: 91-314588/43.

PT New plasmid vector for breeding improved Corynebacterium - contains 3 specified DNA regions for replicating proliferation of improved Corynebacterium

PS Disclosure; Page 4; 13pp; Japanese.

CC This is the sequence of a preferred DNA region (A) containing a gene capable of maintaining stable replicative proliferation of a plasmid in Corynebacterium cells. Plasmids of the invention are composed of such a region (A), along with a region (B) carrying the replicative proliferation function of the plasmid in Corynebacterium and a region (C) containing a drug resistance marker gene. The fragment is excised from plasmid pBV503 which is retained in B. stationis IFO 12144 (FERM BP-2515). A 2.1kb KpnI-EcoRI fragment from the same source can be used for the same function. Region (B) is preferably excised from the same source or from plasmid pBV502 retained in CC B. flavum M233 (FERM BP-1497).

CC Sequence 1763 BP: 433 A: 436 C: 445 G: 449 T:

Query Match 80.0%; Score 80; DB 3; Length 1763;

Best Local Similarity 70.6%; Pred. No. 7.71e+01;
Matches 12; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 1327 acattgcgctttgatg 1343

|||||:|||||:|||||
Cp 23 ACRTTMCRTTYTGRTG 7

RESULT 6

ID 003332 standard; RNA: 1763 BP.

AC 003332.
DT 12-JUL-1990 (first entry)

DE Gene encodes function of maintaining a plasmid stably in a coryne-form bacteria of the genus Brevibacterium.

KW pBV503; cryptophan synthetase; coryneform; ds.
OS Brevibacterium stationis.

PN EP-352763-A.

PD 31-JAN-1990.
PF 26-JUL-1988: 113775.

PR 27-JUL-1988; JP-185428.
PR 08-SEP-1988; JP-223399.

PA (MIRP) Mitsubishi Petroch KK.
PI Kohama K, Kobayashi M, Kurusu Y, Yukawa H;

PT New DNA fragments which stabilise plasmids in coryneform bacteria - and derived recombinant vectors which are retained in cells without selection pressure.

PS Claim 6; Page 23; 27pp; English.

CC Plasmid pBV503 contains DNA fragment present in Brevibacterium stationis

CC IFO 12144, contains gene which stabilises plasmids in the bacteria, allowing maintenance with no selection pressure.

CC Cryptophan synthetase gene with associated promoter/operator system has been maintained within a stable plasmid associated with the gene.

SQ Sequence 1763 BP: 433 A: 436 C: 445 G: 449 T:

Query Match 80.0%; Score 80; DB 1; Length 1763;

Best Local Similarity 70.6%; Pred. No. 7.71e+01;
Matches 12; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 1327 acattgcgctttgatg 1343

|||||:|||||:|||||
Cp 23 ACRTTMCRTTYTGRTG 7

RESULT 7

ID 003742 standard; DNA: 2839 BP.

AC 003742.
DT 15-AUG-1989 (first entry)

DE Human SKI related gene
KW Human sloan Kettering Institute related gene; cancer; ss.
FH Key Location/Qualifiers
FT cds 709..2765

FT /*tag-a

FT /*label-human SKI-related protein

PN J02053485-A.

PD 22-FEB-1990.
PF 19-AUG-1988: 205906.

PA (RIKA) Rikagaku Kenkyusho.
DR WPI: 90-103117-A.

P-PSDB: R03664.

PT Human SKI related gene - coded by base sequence including 709 adenine to 1806 guanine etc.

PS Disclosure; fig 1; 6pp; Japanese.

CC This sequence encodes a human sloan Kettering Institute (SKI) related protein. A truncated protein can also be encoded by bases 709-1806.

CC The protein can be used as a cancer marker by preparing an antibody against it. The copy no. of a cancer gene can be found by Southern

CC blotting using the gene or fragments. The degree of malignancy of a cancer can thus be determined. See also 003743.

CC Sequence 2839 BP: 936 A: 521 C: 586 G: 796 T:

Query Match 80.0%; Score 80; DB 1; Length 2839;

Best Local Similarity 56.5%; Pred. No. 7.71e+01;
Matches 13; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Db 1420 actttcccaaatgtagcgt 1442

|||||:|||||:|||||
QY 1 ACNTMCATCARAATGNAATGT 23

RESULT 8

ID 003743 standard; DNA: 2888 BP.

AC 003743.
DT 15-AUG-1989 (first entry)

DE Human SKI related gene
KW Human sloan Kettering Institute related gene; cancer; ss.
FH Key Location/Qualifiers
FT cds 709..1956

FT /*tag-a

FT /*label-human SKI related protein

PN J02053485-A.

PD 22-FEB-1990.
PF 19-AUG-1988: 205906.

PA (RIKA) Rikagaku Kenkyusho.
DR WPI: 90-103117-A.

P-PSDB: R03665.
PT Human SKI related gene - coded by base sequence including 709 adenine to 1806 guanine etc.
PS Disclosure; fig 2; 6pp; Japanese.
CC This sequence encodes a human sloan Kettering Institute (SKI) related

CC protein. The protein can be used as a cancer marker by preparing an
 CC antibody against it. The copy no. of a cancer gene can be found by
 CC Southern blotting using the gene or fragments. The degree of malignancy
 CC of a cancer can thus be determined. See also 003742.
 SQ Sequence 2888 BP; 857 A; 566 C; 598 G; 867 T;

Query Match 80.0%; Score 80; DB 1; Length 2888;
 Best Local Similarity 56.5%; Pred. No. 7.71e+01;
 Matches 13; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Db 1421 atttctcctcaaatgtagcgt 1443

Oy 1 ACNNTCAYCARAAAYGNAAYGT 23

RESULT 9
 ID T65001 standard; DNA; 5687 BP.

AC T65001;

DT 28-MAY-1997 (first entry)

DE Mouse cell cycle regulatory protein SPA-1 genomic DNA (exons 1-4).

KW Cell division cycle; Span-N; Span-C; lymphocyte differentiation;

OS mouse; murine; ds.

FH Key

FT exon

FT Location/Qualifiers

FT 3109..3284

FT /*tag= a

FT /number= 1

FT exon

FT 3764..4555

FT /*tag= b

FT /number= 2

FT exon

FT 5147..5273

FT /*tag= c

FT /number= 3

FT exon

FT 5383..5524

FT /*tag= d

FT /number= 4

J08217797-A.

PD 27-AUG-1996.

PF 17-FEB-1995; 051999.

PR 30-MAY-1994; JP-139513.

PR 20-OCT-1994; JP-279712.

PR 13-DEC-1994; JP-332520.

PA (MINA/) MINATO N.

DR WPI: 96-439574/44.

PT Gene encoding cell division-regulatory protein, SPA-1 - useful as

PT regulator of cell division and differentiation of lymphocytes

PS Claim 13; Page 12-14; 26pp; Japanese.

CC The present partial genomic DNA sequence comprises exons 1-4 for a

CC gene encoding cell division-regulating protein (designated SPA-1),

CC which is not expressed in the resting phase but is expressed after

CC the growth phase in the cell cycle of mammalian cells. SPA-1 is

CC useful as a regulator of cell division or for differentiation

CC of lymphocytes.

SQ Sequence 5687 BP; 1201 A; 1574 C; 1632 G; 1280 T;

Query Match 80.0%; Score 80; DB 29; Length 5687;

Best Local Similarity 61.1%; Pred. No. 7.71e+01;

Matches 11; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 3448 attccgcttcgtagtaaa 3465

Cp 21 RTTNCCTTGTGTGNAR 4

RESULT 10

ID 068133 standard; DNA; 339 BP.

AC 068133;

DT 26-JUN-1995 (first entry)

DE Sequence encoding human interleukin-3 mutein (pMon13360).

KW Human interleukin-3; hIL-3; mutant; mutein; ss.

OS Synthetic.

PN MO9412638-A.

PD 09-JUN-1994.

PF 22-NOV-1993; U11197.

PR 24-NOV-1992; US-981044.

PA (MONS) MONSANTO CO.

PA (SEAR) SEARLE & CO G D.

PI Abrams MA, Bauer SC, Bradford-Goldberg SR, Caparon MH;

PI Easton AM, Klein BK, McKearn JP, Ollins PO, Palk K;

PI Thomas JW;

DR WPI: 94-200266/24.

DR P-PSDB: R56107.

PT New human interleukin-3 mutant polypeptide(s) - useful for

PT treating various blood cell deficiencies, including leukopenia,

PT neutropenia and leukemia

PS Claim 42; Page 150; 388pp; English.

CC The sequence encodes a mutein of hIL-3, comprising Met-Ala-

CC (15-125)hIL-3. The muteins may have biological activities similar

CC to or better than hIL-3 and, in some cases, may also have an

CC improved side effect profile, e.g. better therapeutic index. They

CC may be used to treat e.g. leukopenia, neutropenia, thrombocytopenia,

CC (aplastic) anemia, Chediak-Higashi syndrome, SLE, leukemia,

CC myelodysplastic syndrome, myelofibrosis; bone marrow suppression or

CC haematopoietic deficiency associated with treatment with AZT or due

CC to dialysis; or immunodeficiency resulting from viral infection,

CC exposure to radiation or cancer treatment.

CC The muteins are made by recombinant DNA techniques. Recombinant DNA

CC sequences encoding the muteins are given in GENSEQ Accession Numbers

SQ Sequence 339 BP; 101 A; 91 C; 68 G; 79 T;

Query Match 75.0%; Score 75; DB 13; Length 339;

Best Local Similarity 63.2%; Pred. No. 2.09e+02;

Matches 12; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 95 acgttctatccgtagga 113

Cp 23 ACRTTNCCTTGTGTGNAR 5

RESULT 11

ID 068132 standard; DNA; 339 BP.

AC 068132;

DE Sequence encoding human interleukin-3 mutein (pMon13475).

DE Human interleukin-3; hIL-3; mutant; mutein; ss.

KW Synthetic.

PN MO9412638-A.

PD 09-JUN-1994.

PF 22-NOV-1993; U11197.

PR 24-NOV-1992; US-981044.

PA (MONS) MONSANTO CO.

PA (SEAR) SEARLE & CO G D.

PI Abrams MA, Bauer SC, Bradford-Goldberg SR, Caparon MH;

PI Easton AM, Klein BK, McKearn JP, Ollins PO, Palk K;

PI Thomas JW;

DR WPI: 94-200266/24.

DR P-PSDB: R56106.

PT New human interleukin-3 mutant polypeptide(s) - useful for

PT treating various blood cell deficiencies, including leukopenia,

PT neutropenia and leukemia

PS Claim 42; Page 150; 388pp; English.

CC The sequence encodes a mutein of hIL-3, comprising Met-Ala-

CC (15-125)hIL-3. The muteins may have biological activities similar

CC to or better than hIL-3 and, in some cases, may also have an

CC improved side effect profile, e.g. better therapeutic index. They

CC may be used to treat e.g. leukopenia, neutropenia, thrombocytopenia,

CC (aplastic) anemia, Chediak-Higashi syndrome, SLE, leukemia,

CC myelodysplastic syndrome, myelofibrosis; bone marrow suppression or

CC haematopoietic deficiency associated with treatment with AZT or due

CC exposure to radiation or cancer treatment.

CC The muteins are made by recombinant DNA techniques. Recombinant DNA

CC sequences encoding the muteins are given in GENSEQ Accession Numbers

SQ Sequence 339 BP; 100 A; 92 C; 70 G; 77 T;

Query Match 75.0%; Score 75; DB 13; Length 339;
 Best Local Similarity 63.2%; Pred. No. 2.09e+02;
 Matches 12; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

DB 95 acgttctatctgtagga 113
 ||:| | :||:| | | |
 CP 23 ACRTTNCRTTYGRTGNA 5

RESULT 12
 ID 068134 standard; DNA; 339 BP.
 AC 068134;
 DT 26-JUN-1995 (first entry)
 DE Sequence encoding human interleukin-3 mutein (pMon13367).
 KW Human Interleukin-3; hIL-3; mutant; mutein; ss.
 OS Synthetic.
 PN W09412638-A.
 PD 09-JUN-1994.
 PF 22-NOV-1993; U11197.
 PR 24-NOV-1992; US-981044.
 PA (MONS) MONSANTO CO.
 PI (SEAR) SEARLE & CO G D.
 PI Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH;
 PI Easton AM, Klein BK, McKearn JP, Ollins PO, Paik K;
 PI Thomas JW;
 DR WPI; 94-200266/24.
 DR P-PSDB; R56108.
 PT New human interleukin-3 mutant polypeptide(s) - useful for
 PT treating various blood cell deficiencies, including leukopenia,
 PT neutropenia and leukaemia.
 PS Claim 42; Page 150; 388pp; English.
 CC The sequence encodes a mutein of hIL-3, comprising Met-Ala-
 CC (15-125)hIL-3. The muteins may have biological activities similar
 CC to or better than hIL-3 and, in some cases, may also have an
 CC improved side effect profile, e.g. better therapeutic index. They
 CC may be used to treat e.g. leukopenia, neutropenia, thrombocytopenia,
 CC (aplastic) anemia, Chediak-Higashi syndrome, SLE, leukaemia,
 CC myelodysplastic syndrome, myelofibrosis; bone marrow suppression or
 CC haematopoietic deficiency associated with treatment with AZT or due
 CC to dialysis; or immunodeficiency resulting from viral infection,
 CC exposure to radiation or cancer treatment.
 CC The muteins are made by recombinant DNA techniques. Recombinant DNA
 CC sequences encoding the muteins are given in GENESEQ Accession Numbers
 CC 068083 - 068189.

SQ Sequence 339 BP; 99 A; 91 C; 69 G; 80 T;

Query Match 75.0%; Score 75; DB 13; Length 339;
 Best Local Similarity 63.2%; Pred. No. 2.09e+02;
 Matches 12; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

DB 95 acgttctatctgtagga 113
 ||:| | :||:| | | |
 CP 23 ACRTTNCRTTYGRTGNA 5

RESULT 13
 ID 068135 standard; DNA; 339 BP.
 AC 068135;
 DT 26-JUN-1995 (first entry)
 DE Sequence encoding human interleukin-3 mutein (pMon13369).
 KW Human Interleukin-3; hIL-3; mutant; mutein; ss.
 OS Synthetic.
 PN W09412638-A.
 PD 09-JUN-1994.
 PF 22-NOV-1993; U11197.
 PR 24-NOV-1992; US-981044.
 PA (MONS) MONSANTO CO.
 PI (SEAR) SEARLE & CO G D.
 PI Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH;
 PI Easton AM, Klein BK, McKearn JP, Ollins PO, Paik K;
 PI Thomas JW;
 DR WPI; 94-200266/24.

DR P-PSDB; R56109.
 PT New human interleukin-3 mutant polypeptide(s) - useful for
 PT treating various blood cell deficiencies, including leukopenia,
 PT neutropenia and leukaemia
 PS Claim 42; Page 150-151; 388pp; English.
 CC The sequence encodes a mutein of hIL-3, comprising Met-Ala-
 CC (15-125)hIL-3. The muteins may have biological activities similar
 CC to or better than hIL-3 and, in some cases, may also have an
 CC improved side effect profile, e.g. better therapeutic index. They
 CC may be used to treat e.g. leukopenia, neutropenia, thrombocytopenia,
 CC (aplastic) anemia, Chediak-Higashi syndrome, SLE, leukaemia,
 CC myelodysplastic syndrome, myelofibrosis; bone marrow suppression or
 CC haematopoietic deficiency associated with treatment with AZT or due
 CC to dialysis; or immunodeficiency resulting from viral infection,
 CC exposure to radiation or cancer treatment.
 CC The muteins are made by recombinant DNA techniques. Recombinant DNA
 CC sequences encoding the muteins are given in GENESEQ Accession Numbers
 CC 068083 - 068189.

SQ Sequence 339 BP; 102 A; 93 C; 66 G; 78 T;

Query Match 75.0%; Score 75; DB 13; Length 339;
 Best Local Similarity 54.5%; Pred. No. 2.09e+02;
 Matches 12; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

DB 95 acatgacatctgtaggac 116
 ||:| | :||:| | | |
 CP 23 ACRTTNCRTTYGRTGNANG 2

RESULT 14
 ID 068115 standard; DNA; 339 BP.
 AC 068115;
 DT 08-AUG-1995 (first entry)
 DE Sequence encoding human interleukin-3 mutein (pMon13300).
 KW Human Interleukin-3; hIL-3; mutant; mutein; ss.
 OS Synthetic.
 PN W09412638-A.
 PD 09-JUN-1994.
 PF 22-NOV-1993; U11197.
 PR 24-NOV-1992; US-981044.
 PA (MONS) MONSANTO CO.
 PI (SEAR) SEARLE & CO G D.
 PI Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH;
 PI Easton AM, Klein BK, McKearn JP, Ollins PO, Paik K;
 PI Thomas JW;
 DR WPI; 94-200266/24.
 PT New human interleukin-3 mutant polypeptide(s) - useful for
 PT treating various blood cell deficiencies, including leukopenia,
 PT neutropenia and leukaemia
 PS Claim 42; Page 145-146; 388pp; English.
 CC The sequence encodes a mutein of hIL-3, comprising Met-Ala-
 CC (15-125)hIL-3. The muteins may have biological activities similar
 CC to or better than hIL-3 and, in some cases, may also have an
 CC improved side effect profile, e.g. better therapeutic index. They
 CC may be used to treat e.g. leukopenia, neutropenia, thrombocytopenia,
 CC (aplastic) anemia, Chediak-Higashi syndrome, SLE, leukaemia,
 CC myelodysplastic syndrome, myelofibrosis; bone marrow suppression or
 CC haematopoietic deficiency associated with treatment with AZT or due
 CC to dialysis; or immunodeficiency resulting from viral infection,
 CC exposure to radiation or cancer treatment.
 CC The muteins are made by recombinant DNA techniques. Recombinant DNA
 CC sequences encoding the muteins are given in GENESEQ Accession Numbers
 CC 068083 - 068189.


```

RESULT 15
ID 068146 standard; DNA; 339 BP.
AC 068146;
DT 26-JUN-1995 (first entry)
DE Sequence encoding human interleukin-3 mutein (pMon13383).
KW Human interleukin-3; hIL-3; mutant; mutein; ss.
OS Synthetic.
PN MO9412638-A.
PD 09-JUN-1994.
PF 22-NOV-1993; U11197.
PR 24-NOV-1992; US-981044.
PA (MONS ) MONSANTO CO.
PB (SEAR ) SEARLE & CO G D.
PI Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH;
PI Easton AM, Klein BK, McKearn JP, Olin PO, Paik K;
PI Thomas JW;
PI WPI; 94-200266/24.
DR P-PSDB; R56120.
PT New human interleukin-3 mutant polypeptide(s) - useful for
PT treating various blood cell deficiencies, including leukopenia,
PT neutropenia and leukaemia
PS Claim 42; Page 153; 388pp; English.
CC The sequence encodes a mutein of hIL-3, comprising Met-Ala-
CC (15-125)hIL-3. The muteins may have biological activities similar
CC to or better than hIL-3 and, in some cases, may also have an
CC improved side effect profile, e.g. better therapeutic index. They
CC may be used to treat e.g. leukopenia, neutropenia, thrombocytopenia,
CC (aplastic) anemia, Chediak-Higashi syndrome, SLE, leukaemia,
CC myelodysplastic syndrome, myelofibrosis; bone marrow suppression or
CC haematopoietic deficiency associated with treatment with AZT or due
CC to dialysis; or immunodeficiency resulting from viral infection,
CC exposure to radiation or cancer treatment.
CC The muteins are made by recombinant DNA techniques. Recombinant DNA
CC sequences encoding the muteins are given in GENESEQ Accession Numbers
CC 068083 - 068189.
CC Sequence 339 BP; 101 A; 92 C; 69 G; 77 T;

Query Match 75.0%; Score 75; DB 13; Length 339;
Best Local Similarity 63.2%; Pred. No. 2.09e+02;
Matches 12; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 95 acgttcattctgataaga 113
   ||:|:| | :||:|:|:| |
CP 23 ACRTTNCRTTYTGRTGNA 5

```

Search completed: Tue Apr 27 10:35:39 1999
 Job time : 28 secs.

(TM)

Distribution rights by Oxford Molecular Ltd

Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Apr 27 10:38:46 1999; Maspar time 1498.14 Seconds

Tabular output not generated.

Title: >US-09-011-797-1

N.A. Sequence: 1 CTCTCCAGCGTGTTCAGCAG.....AAAATAAAGGATTC 932

Comp: GAGAGGTCGCACAGTCGTC.....TTTTTTTTTTTCCCTTAAG

Scoring table: TABLE default

Nmatch STD : Dbase 0; Query 0

Searched: 602357 seqs, 1181590623 bases x 2

Post-processing: Minimum Match 08

Database:

1:em_ba 2:em_fun 3:em_htg 4:em_huml 5:em_humz 6:em_pi
7:em_o 8:em_or 9:em_u 10:em_pat 11:em_sts 12:em_syn
13:em_xo
embanK10
12:gb_ba1 17:gb_ba2 18:gb_htg_in 20:gb_o 21:gb_ov
22:gb_pat 23:gb_p12 24:gb_p11 25:gb_p12 26:gb_p1
27:gb_p2 28:gb_p3 29:gb_xo 30:gb_st 31:gb_sts 32:gb_sy
33:gb_un 34:gb_v

Statistics: Mean 11.237; Variance 7.557; scale 1.487

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	923	99.0	925	29	579730	ORL1 receptor agonist	0.00e+00
2	923	99.0	988	29	RNPNNRNA	R.norvegicus mRNA for	0.00e+00
3	904	97.0	972	29	RUN48262	Rattus norvegicus pre-	0.00e+00
4	554	59.4	1354	29	MUSNOP	Mouse mRNA for nocleop	0.00e+00
5	550	59.0	1047	29	MUSN2K	Mouse N23K mRNA for N2	0.00e+00
6	366	39.3	633	29	MPENNRNA	M.musculus mRNA for pr	5.37e-19
7	355	38.1	972	29	MUSN27K	Mouse N23K mRNA for d	4.23e-19
8	336	36.1	1015	26	HSPNNRNA	H.septiens mRNA for pre	3.33e-18
9	336	36.1	1198	28	HSU48263	Human pre-pro-orphalin	3.33e-18
10	299	32.1	1253	29	MPEPNX3	M.musculus gene encodi	4.21e-15
11	296	31.8	1209	20	ABO05521	Bos taurus mRNA for no	3.11e-15
12	236	25.3	372	29	MMU40427	Mus musculus orphalin	5.03e-11
13	228	24.5	878	26	HSPPNX3	H.septiens gene encodin	4.30e-11

14	157	16.8	670	26	HSPNEX4	H.sapiens gene encod	8.96e-7	
C	15	131	14.1	364	31	human STS Wt-15701	3.27e-55	
16	87	9.3	557	29	MMPNEX2	R.novgorgius gene enco	1.44e-33	
17	71	7.6	400	29	MMPNEX2	M.musculus gene encod	1.31e-22	
C	18	7.1	7218	22	IM6494	Sequence 14 from paten	7.16e-11	
19	66	6.4	7218	22	IM6494	Sequence 14 from paten	1.12e-11	
C	20	6.0	553	26	HSPNEX2	H.sapiens gene encod	1.61e-11	
C	21	4.1	965	22	AR024229	Sequence 22 from paten	1.47e-07	
C	22	3.6	74311	27	AC005369	Home sapiens chromosom	4.42e-03	
C	23	3.5	10772	19	AF012089	Drosophila melanogaste	3.35e-00	
C	24	3.5	10772	19	AF012089	Drosophila melanogaste	9.27e-07	
C	25	3.4	215	22	128278	Sequence 5 from patent	9.27e-07	
C	26	3.4	215	22	128278	Sequence 5 from patent	9.27e-07	
C	27	3.3	965	22	AR024229	Sequence 22 from paten	2.53e-07	
C	28	3.3	74311	27	AC005369	Home sapiens chromosom	2.53e-07	
C	29	3.2	181992	18	AC003066	*** SEQUENCING IN PROG	6.79e-07	
C	30	3.2	186314	29	AC003694	Mus musculus chromosom	6.79e-07	
C	31	3.2	186314	29	AC003694	Mus musculus chromosom	6.79e-07	
C	32	2.9	3.1	2691	29	RATCK1	Rat gene for choline x	1.79e+00
C	33	2.9	3.1	12387	29	AF03857	Mus musculus NSP-like	1.79e+00
C	34	2.9	3.1	11244	18	AC003022	*** SEQUENCING IN PROG	1.79e+00
C	35	2.9	3.1	119101	29	MMAE00665	Mus musculus TCR beta	1.79e+00
C	36	2.9	3.1	216021	27	H0AC004787	Home sapiens Chromosom	1.79e+00
C	37	2.8	3.0	884	29	MMAJ2012	Mus musculus DNA for B	4.66e+00
C	38	2.8	3.0	1508	27	AF052142	Home sapiens clone 246	4.66e+00
C	39	2.8	3.0	1775	24	SCUBI2	Yeast UB12 gene for ub	4.66e+00
C	40	2.8	3.0	9424	29	MUSLFR02	Murine lactoferrin gen	4.66e+00
C	41	2.8	3.0	10288	29	MMA01033	Mus musculus kcc2 gene	4.66e+00
C	42	2.8	3.0	11934	29	MMA6033	Mus musculus c15k gene	4.66e+00
C	43	2.8	3.0	43601	29	MMA03361	M.musculus COL3A1 gene	4.66e+00
C	44	2.8	3.0	108400	28	HUMDC028	Home sapiens Digeorge	4.66e+00
C	45	2.8	3.0	114617	24	AT13C5	Arabidopsis thaliana D	4.66e+00

RESULT	1	30-JAN-1996
LOCUS	579730	925 bp mRNA
DEFINITION	ORL1 receptor agonist precursor-endogenous agonist of opioid receptor-like ORL1 receptor [rats, brain, mRNA Partial, 925 nt].	
ACCESSION	579730	
NID	g1172238	
KEYWORDS		
SOURCE	Rattus sp. brain.	
ORGANISM	Rattus sp.	
REFERENCE	Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
AUTHORS	1 (bases 1 to 925) Meunier,J.C., Mollereau,C., Toll,L., Snaudreau,C., Moisan,C., Alvinerie,P., Butour,J.L., Guillemot,J.C., Ferrara,P., Monsarrat,B. et.al.	
TITLE	Isolation and structure of the endogenous agonist of opioid receptor-like ORL1 receptor [see comments]	
JOURNAL	Nature 377 (6549), 532-535 (1995)	
MEDLINE	96013745	
REMARK	GenBank staff at the National Library of Medicine created this entry [NCBI g1bseq 171655] from the original journal article. This sequence comes from Fig. 4.	
FEATURES	Location/Qualifiers	
SOURCE	1..925	
	/organism="Rattus sp."	
	/db_xref="taxon:10118"	
gene	1..507	
	/partial	
	/gene="ORL1 receptor agonist precursor"	
CDS	1..507	
	/partial	
	/gene="ORL1 receptor agonist precursor"	
	/note="endogenous agonist of opioid receptor-like ORL1 receptor; Authors indicate ORL1 receptor agonist peptide is encoded by residues 122-138. Method: conceptual translation with partial peptide sequencing. This sequence	


```

comes from Fig. 4"
/codon_start=1
/product="ORF1 receptor agonist precursor"
/db_xref="PID:g1172239"
/translation="LSSVSSSCPEDCLTCOERLHPAPSEFNKILICCEKVPRL
WLTCTAMASDSOLSPADPELTSAIYOSKASEMHRKRVSVYQARAEAD
AEVDADEADVEOKOLKRFSGFTGARKARLANOKRSEFMRQIYLVSMOSSQRR
TLHONGNV"
BASE COUNT      222 a      271 c      237 g      195 t
ORIGIN

Query Match      99.0%; Score 923; DB 29; Length 925;
Best Local Similarity 99.9%; Pred. No. 0.00e+00;
Matches 924; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db      1 CTCCTCAGCGTGTTCAGACCTGTCGCCGAGAGATGCTCTACCTGCCAGAGAGGCTCCAC 60
QY      1 CTCCTCAGCGTGTTCAGACCTGTCGCCGAGAGATGCTCTACCTGCCAGAGAGGCTCCAC 60
QY      1 CTCCTCAGCGTGTTCAGACCTGTCGCCGAGAGATGCTCTACCTGCCAGAGAGGCTCCAC 60
Db      61 CCGGCTCCGGGAGCTTCAACCTGAAGCTGTGCATCTCCAGTGTGAAGAGAGGCTTC 120
QY      61 CCGGCTCCGGGAGCTTCAACCTGAAGCTGTGCATCTCCAGTGTGAAGAGAGGCTTC 120
QY      61 CCGGCTCCGGGAGCTTCAACCTGAAGCTGTGCATCTCCAGTGTGAAGAGAGGCTTC 120
Db      121 CCGGCTCCGGGAGCTTCAACCTGAAGCTGTGCATCTCCAGTGTGAAGAGAGGCTTC 120
QY      121 CCGGCTCCGGGAGCTTCAACCTGAAGCTGTGCATCTCCAGTGTGAAGAGAGGCTTC 120
QY      121 CCGGCTCCGGGAGCTTCAACCTGAAGCTGTGCATCTCCAGTGTGAAGAGAGGCTTC 120
Db      181 CCTGCTGATCCAGAGCTCAGCTCCGCTCTTACAGTGGAAAGCCTCGGAGATGAC 240
QY      181 CCTGCTGATCCAGAGCTCAGCTCCGCTCTTACAGTGGAAAGCCTCGGAGATGAC 240
QY      181 CCTGCTGATCCAGAGCTCAGCTCCGCTCTTACAGTGGAAAGCCTCGGAGATGAC 240
Db      241 CACCTGAAGAGATGCCGCGTCTCAGAGTGTGTGCAAGCCGAGAGAGAGAGGCTGAG 300
QY      241 CACCTGAAGAGATGCCGCGTCTCAGAGTGTGTGCAAGCCGAGAGAGAGAGGCTGAG 300
QY      241 CACCTGAAGAGATGCCGCGTCTCAGAGTGTGTGCAAGCCGAGAGAGAGAGGCTGAG 300
Db      301 GCAGATGCAGAGCTGTGCGAGATAGGCGCATGAGTGGAGAGAGAGAGAGAGGCTGAC 360
QY      301 GCAGATGCAGAGCTGTGCGAGATAGGCGCATGAGTGGAGAGAGAGAGAGAGGCTGAC 360
QY      301 GCAGATGCAGAGCTGTGCGAGATAGGCGCATGAGTGGAGAGAGAGAGAGAGGCTGAC 360
Db      361 AGGTTGGGGGCTTCTACTGGGGCCCGGAAGTCAAGCCGGAAGTGGGCAACGAAAGCGG 420
QY      361 AGGTTGGGGGCTTCTACTGGGGCCCGGAAGTCAAGCCGGAAGTGGGCAACGAAAGCGG 420
QY      361 AGGTTGGGGGCTTCTACTGGGGCCCGGAAGTCAAGCCGGAAGTGGGCAACGAAAGCGG 420
Db      421 TTCAGTGAATTTATGAGGAGTACCTGCTGTCAGATGCAAGTCAAGCCGAGAGG 480
QY      421 TTCAGTGAATTTATGAGGAGTACCTGCTGTCAGATGCAAGTCAAGCCGAGAGG 480
QY      421 TTCAGTGAATTTATGAGGAGTACCTGCTGTCAGATGCAAGTCAAGCCGAGAGG 480
Db      481 ACTCTGCACCAAGATGTGTAAATGTGTAGCCAGAAAGAGAGAGAGAGAGAGAGG 540
QY      481 ACTCTGCACCAAGATGTGTAAATGTGTAGCCAGAAAGAGAGAGAGAGAGAGAGG 540
QY      481 ACTCTGCACCAAGATGTGTAAATGTGTAGCCAGAAAGAGAGAGAGAGAGAGAGG 540
Db      541 TGCNACCATGAGATGAGTACAGTCCCGCTGACAGATGTGTGCAACCAAGAGAGG 600
QY      541 TGCNACCATGAGATGAGTACAGTCCCGCTGACAGATGTGTGCAACCAAGAGAGG 600
QY      541 TGCNACCATGAGATGAGTACAGTCCCGCTGACAGATGTGTGCAACCAAGAGAGG 600
Db      601 GCGCGGAGTCAAGATGCTCTCTCCCTGAGGAGTGAACACCCCGGAGAGAGG 660
QY      601 GCGCGGAGTCAAGATGCTCTCTCCCTGAGGAGTGAACACCCCGGAGAGAGG 660
QY      601 GCGCGGAGTCAAGATGCTCTCTCCCTGAGGAGTGAACACCCCGGAGAGAGG 660
Db      661 GCATGCTCACCACCAATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
QY      661 GCATGCTCACCACCAATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
QY      661 GCATGCTCACCACCAATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
Db      721 ACATTTTAAATGAGCCCATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
QY      721 ACATTTTAAATGAGCCCATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
QY      721 ACATTTTAAATGAGCCCATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
Db      781 GAACATGAATAGACCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
QY      781 GAACATGAATAGACCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
QY      781 GAACATGAATAGACCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
Db      841 CTGTCTTTTACATGATGTGTGAACAATTTTCCATTAACTTACTAGACAAA 900
QY      841 CTGTCTTTTACATGATGTGTGAACAATTTTCCATTAACTTACTAGACAAA 900
QY      841 CTGTCTTTTACATGATGTGTGAACAATTTTCCATTAACTTACTAGACAAA 900

```

```

QY      841 CTGTCTTTTACATGATGTGTGAACAATTTTCCATTAACTTACTAGACAAA 900
Db      901 TGGTTAATATAAAAAAAAAAAAAAAAAA 925
QY      901 TGGTTAATATAAAAAAAAAAAAAAAAAA 925
QY      901 TGGTTAATATAAAAAAAAAAAAAAAAAA 925

RESULT      2
LOCUS      RNPmRNA      988 bp      RNA
DEFINITION      R.norvegicus mRNA for prepronociceptin.
ACCESSION      X97375
NID      91532064
KEYWORDS      prepronociceptin.
SOURCE      Norway rat.
ORGANISM      Rattus norvegicus
Eukaryote; Mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE      1 (bases 1 to 988)
AUTHORS      Mollereau,C., Simons,M.J., Soulaire,P., Liners,F., Vassart,G.,
Meunier,J.C. and Parmentier,M.
TITLE      Structure, tissue distribution, and chromosomal localization of the
prepronociceptin gene
JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 93 (16), 8666-8670 (1996)
REFERENCE      2 (bases 1 to 988)
AUTHORS      Parmentier,M.
TITLE      Direct Submission
JOURNAL      Submitted (19-APR-1996) M. Parmentier, Universite Libre de
Bruxelles, 1 R I B N ULB Campus Erasme, 808 Route de Lennik, B-
1070 Bruxelles, BELGIUM
COMMENT      Related sequence 579730.
FEATURES      Location/Qualifiers
source      1..988
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone_1b="lambda DASHII (Stratagene #943501)"
/clone_1b="lambda ZAPRI (brain cDNA)"
/clone="RGEN109"
/clone="RB101"
<1..150
/number=2
join<1..150,151..615,616..974)
25..69
join(25..150,151..570)
/codon_start=1
/product="prepronociceptin"
/db_xref="PID:g1532065"
/translation="KILFCVLTLLSLSVSSCPEDCLTCOERLHPAPSEFNKILIC
EKEVPEPRLWLTCTKAMASDSOLSPADPELTSAIYOSKASEMHRKRVSVYQARAEAD
AEVDADEADVEADVEOKOLKRFSGFTGARKARLANOKRSEFMRQIYLVSMOSSQRR
TYLVSMOSSQRRTLHONGNV"
70..567
/product="prepronociceptin"
151..615
/number=3
616..974
/number=4
971
BASE COUNT      231 a      292 c      250 g      215 t
ORIGIN

Query Match      99.0%; Score 923; DB 29; Length 988;
Best Local Similarity 99.9%; Pred. No. 0.00e+00;
Matches 924; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db      64 CTCCTCAGCGTGTTCAGAGGTCGCCGAGAGATGCTCTACCTGCCAGAGAGGCTCCAC 123
QY      1 CTCCTCAGCGTGTTCAGAGGTCGCCGAGAGATGCTCTACCTGCCAGAGAGGCTCCAC 123
QY      1 CTCCTCAGCGTGTTCAGAGGTCGCCGAGAGATGCTCTACCTGCCAGAGAGGCTCCAC 123
Db      124 CCGGCTCCGGGAGCTTCAACCTGAAGCTGTGCATCTCCAGTGTGAAGAGAGGCTTC 183
QY      124 CCGGCTCCGGGAGCTTCAACCTGAAGCTGTGCATCTCCAGTGTGAAGAGAGGCTTC 183
QY      124 CCGGCTCCGGGAGCTTCAACCTGAAGCTGTGCATCTCCAGTGTGAAGAGAGGCTTC 183

```


QY 61 CCGGCTCCGGGACGCTTCAACCTGAAGCTGATCTCTCAAGTGTGAAGAGAGCTTTC 120
 Db 184 CCGGCTCTCTCTGAGCTCTTTGACCAAGCCATGCGCAGTGTGAGTCTGAGCAGCTCAGC 243
 QY 121 CCGGCTCTCTCTGAGCTCTTTGACCAAGCCATGCGCAGTGTGAGTCTGAGCAGCTCAGC 180
 Db 244 CCGTGTATGACAGAGCTGAGCTGCGCTCTCTTTACAGTGTGAAGCCCTGGAGATGAG 303
 QY 181 CCGTGTATGACAGAGCTGAGCTGCGCTCTCTTTACAGTGTGAAGCCCTGGAGATGAG 240
 Db 304 CACCTGAAGAGAGATGCGCGTGTGAGAGTGTGTGCAAGCCCGAGAGCAGAGCTGAG 363
 QY 241 CACCTGAAGAGAGATGCGCGTGTGAGAGTGTGTGCAAGCCCGAGAGCAGAGCTGAG 300
 Db 364 GCAGTGTGAAGAGCTGTGTGAGAGTGTGTGCAAGCCCGAGAGCAGAGCTGAG 423
 QY 301 GCAGTGTGAAGAGCTGTGTGAGAGTGTGTGCAAGCCCGAGAGCAGAGCTGAG 360
 Db 424 AGGTTGGGGGCTTCACTGGGGCCGGAAGTCAAGCCCGGAAGTTGGCCAGCAGAGCGG 483
 QY 361 AGGTTGGGGGCTTCACTGGGGCCGGAAGTCAAGCCCGGAAGTTGGCCAGCAGAGCGG 420
 Db 484 TTCAAGTGTATGAGAGCAGTACCTGTCTCTGAGCAGTCAAGTCAAGCCAGCAGCGGCGC 543
 QY 421 TTCAAGTGTATGAGAGCAGTACCTGTCTCTGAGCAGTCAAGTCAAGCCAGCAGCGGCGC 480
 Db 544 ACTGTGACACCAAGATGTGTAATGTGTAGCCAGAGAGCCCTCCAGCTGACAGCGGCGAC 603
 QY 481 ACTGTGACACCAAGATGTGTAATGTGTAGCCAGAGAGCCCTCCAGCTGACAGCGGCGAC 540
 Db 604 TGCACCCATGAGCAGTCAAGTGTGAGCCCGGAGAGTGTGTCACAGCCAGAGAGCTGCA 663
 QY 541 TGCACCCATGAGCAGTCAAGTGTGAGCCCGGAGAGTGTGTCACAGCCAGAGAGCTGCA 600
 Db 664 GCGCGGAGTCAAGATCTCTCTCTCTGAGGAGTCAAGACCCGCGAGCTCTCCAGCA 723
 QY 601 GCGCGGAGTCAAGATCTCTCTCTCTGAGGAGTCAAGACCCGCGAGCTCTCCAGCA 660
 Db 724 GCATGTCTCACCACATCTCTGTCTACATCAGAGTGTATTTTGTAACTCTCCAGCTA 783
 QY 661 GCATGTCTCACCACATCTCTGTCTACATCAGAGTGTATTTTGTAACTCTCCAGCTA 720
 Db 784 ACATTTAATGGGCCCCATCTTCTGTCATCTCTGCGCTCTCGAGGCGCAGGTGAGAG 843
 QY 721 ACATTTAATGGGCCCCATCTTCTGTCATCTCTGCGCTCTCGAGGCGCAGGTGAGAG 780
 Db 844 GAACATGAATCAGACCTGCGGCTTTGCTCAGCAGTCTGACATAGTGTGTGTAAAGGAG 903
 QY 781 GAACATGAATCAGACCTGCGGCTTTGCTCAGCAGTCTGACATAGTGTGTGTAAAGGAG 840
 Db 904 CTGTTCTTTTGTGAGTGTGTGTGAAGAACTTCTCATTAAGTCTTACTGAGCAAAA 963
 QY 841 CTGTTCTTTTGTGAGTGTGTGTGAAGAACTTCTCATTAAGTCTTACTGAGCAAAA 900
 Db 964 TGGTTAATAAAAAAGAAAAA 988
 QY 901 TGGTTAATAAAAAAGAAAAA 925
 RESULT 3 RNU48262 972 bp mRNA ROD 23 AUG-1996
 LOCUS Rattus norvegicus pre-pro-orphanin FQ mRNA, complete cds.
 DEFINITION U48262
 ACCESSION U48262
 NID 91185011
 KEYWORDS Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
 Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
 Murinae; Rattus.
 REFERENCE 1 (bases 1 to 972)
 AUTHORS Nothacker, H.P., Reinscheid, R.K., Mansour, A., Henningsen, R.A.,
 Arendt, A., Monsma, F.J. Jr., Watson, S.J. and Clivell, O.

TITLE Primary structure and tissue distribution of the orphanin FQ
 JOURNAL Precursor
 MEDLINE Proc. Natl. Acad. Sci. U.S.A. 93 (16), 8677-8682 (1996)
 REFERENCE 2 (bases 1 to 972)
 AUTHORS Nothacker, H.-P. and Henningsen, R.A.
 TITLE Direct Submission
 JOURNAL Submitted (02-FEB-1996) Hans-Peter Nothacker, PRN 69/202,
 Hoffmann-La Roche AG, Grenzacherstr 124, Basel, 4070, Switzerland
 FEATURES
 source
 1..972
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 CDS
 28..84
 /product="pre-pro-orphanin FQ"
 /codon_start=1
 /translation="MKILFDVLLSLSSVFSSEPCEDLTCQERLHPAPSENYLKIC
 ILCOEIVFPRPMTLCTKMAASDSBOLSPADPELTSAALYSKASAMORLKRPRVR
 SVQARDAEPADAEVADAEVDEKQLOKRFGEFTGARKSARKLANORRFESEFRQ
 YLVLSKQSSQRRRLHQNHY"
 85..570
 /product="pro-orphanin FQ"
 574..972
 3'UTR
 972..972
 polyA_site
 /note="45 A nucleotides"
 BASE COUNT 214 a 293 c 251 g 214 t
 ORIGIN
 Query Match 97.0%; Score 904; DB 29; Length 972;
 Best Local Similarity 99.9%; Pred. NO. 0.00e+00;
 Matches 905; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Db 67 CTCCTCAGCGTGTGACAGCTGTGCTCCGAGAGTGTGCTTCACTGCGCAGAGAGCTCCAC 126
 QY 1 CTCCTCAGCGTGTGACAGCTGTGCTCCGAGAGTGTGCTTCACTGCGCAGAGAGCTCCAC 60
 Db 127 CCGGCTCCGGGACGCTTCAACCTGAAGTGTGCACTCTCAGTGTGAAGAGAGTCTTC 186
 QY 61 CCGGCTCCGGGACGCTTCAACCTGAAGTGTGCACTCTCAGTGTGAAGAGAGTCTTC 120
 Db 187 CCGGCTCTCTCTGAGCTCTTTGACCAAGCCATGCGCAGTGTGAGCAGCTCAGC 246
 QY 121 CCGGCTCTCTCTGAGCTCTTTGACCAAGCCATGCGCAGTGTGAGCAGCTCAGC 180
 Db 247 CCGTGTATGACAGAGCTGAGCTGCGCTCTCTTACAGAGTGTGAAGAGCTCGAGATGAG 306
 QY 181 CCGTGTATGACAGAGCTGAGCTGCGCTCTCTTACAGAGTGTGTGAAGAGCTCGAGATGAG 240
 Db 307 CACCTGAAGAGAGATGCGCGTGTGAGAGTGTGTGAAGCCCGAGAGCAGAGCTGAG 366
 QY 241 CACCTGAAGAGAGATGCGCGTGTGAGAGTGTGTGAAGCCCGAGAGCAGAGCTGAG 300
 Db 367 GCAGATGAGAGAGCTGTGTGAGAGTGTGTGAAGCCCGAGAGCAGAGCTGAG 426
 QY 301 GCAGATGAGAGAGCTGTGTGAGAGTGTGTGAAGCCCGAGAGCAGAGCTGAG 360
 Db 427 AGGTTGGGGGCTTCACTGGGGCCGGAAGTCAAGCCCGGAAGTTGGCCAGCAGAGCGG 486
 QY 361 AGGTTGGGGGCTTCACTGGGGCCGGAAGTCAAGCCCGGAAGTTGGCCAGCAGAGCGG 420
 Db 487 TTCAAGTGTATGAGAGCAGTACCTGTCTCTGAGCAGTCAAGCAGCAGCGGCGC 546
 QY 421 TTCAAGTGTATGAGAGCAGTACCTGTCTCTGAGCAGTCAAGCAGCAGCGGCGC 480
 Db 547 ACTGTGACACCAAGATGTGTAATGTGTAGCCAGAGAGCCCTCCAGCTGACAGCGGCGAC 606
 QY 481 ACTGTGACACCAAGATGTGTAATGTGTAGCCAGAGAGCCCTCCAGCTGACAGCGGCGAC 540
 Db 607 TGCACCCATGAGCAGTCAAGTGTGAGCCCGGAGAGTGTGTCACAGCCAGAGAGCTTCA 666

QY 541 TGCAACCATGAGCATCCAGGTGAGCCGCCCTACAGCATGTGTCCACACCAAGACTGCA 600
 Db 667 GGGCGGAGTGCAGATTCCTCTCTCCCTGAGGACAGTAAACCCGCGGACCTCCACCA 726
 QY 601 GGGCGGAGTGCAGATTCCTCTCTCCCTGAGGACAGTAAACCCGCGGACCTCCACCA 660
 Db 727 GCATGTCTCAGCAATCTCTGTCTGATCATCATGAGTATTTTGTAAATCTCCAGCTA 786
 QY 661 GCATGTCTCAGCAATCTCTGTCTGATCATCATGAGTATTTTGTAAATCTCCAGCTA 720
 Db 787 ACATTTTAATGCCCCATCTTCTGTCATCTCTGCTCTCTGAGGCGGAGTGAAGAG 846
 QY 721 ACATTTTAATGCCCCATCTTCTGTCATCTCTGCTCTCTGAGGCGGAGTGAAGAG 780
 Db 847 GAACATGAATTCAGACTGGGGTTTGGCCCTCAGCAGCTAGCTAGTGTGAAGAGAG 906
 QY 781 GAACATGAATTCAGACTGGGGTTTGGCCCTCAGCAGCTAGCTAGTGTGAAGAGAG 840
 Db 907 CTGTCTTTTGAAGTGTGTTGAAACAACTTCTCATTAACCTTCTAGTGAAGAGAG 966
 QY 841 CTGTCTTTTGAAGTGTGTTGAAACAACTTCTCATTAACCTTCTAGTGAAGAGAG 900
 Db 967 TGGTTA 972
 QY 901 TGGTTA 906

RESULT 4
 LOCUS MUSNOP 1354 bp mRNA
 DEFINITION Mouse mRNA for nociceptin/orphanin FQ, complete cds.
 ACCESSION D82866
 NID 91311472
 KEYWORDS nociceptin/orphanin FQ.
 SOURCE Mus musculus adult brain cDNA to mRNA.
 ORGANISM Mus musculus
 Eukaryotes; mitochondria; eukaryotes; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 1354)
 Takeshima, H.
 Direct Submission
 Submitted (25-DEC-1995) to the DDBJ/EMBL/GenBank databases. Hiroshi
 Takeshima, University of Tokyo, Dept. of Pharmacology, Faculty of
 Medicine, Hongo 7-3-1, Bunkyo-ku, Tokyo 113, Japan
 (E-mail: takeshimem.u-tokyo.ac.jp, Tel: 03-3812-2111 (ex. 3422),
 Fax: 03-3815-9360)
 2 (bases 1 to 1354)
 Takeshima, H.
 Unpublished (1996)
 3 (sites)
 Houtani, T., Nishi, M., Takeshima, H., Nukada, T. and Sugimoto, T.
 Structure and regional distribution of nociceptin/orphanin FQ
 precursor
 Biochem. Biophys. Res. Commun. 219 (3), 714-719 (1996)
 96216718
 JOURNAL MEDLINE
 FEATURES
 source

Location/Qualifiers
 1..1354
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /dev_stage="adult"
 /tissue_type="brain"
 360..923
 /codon_start=1
 /product="nociceptin/orphanin FQ"
 /db_xref="PID:d101281"
 /db_xref="PID:g1311473"
 /translation="MKILFCDVLLSLSSVSSPRLCTCOELHPADSFNKTCT
 ILQCEKVPRLMTCTKVMASGSGQLSPADPEVSAALVTPKASEQHLKMPVR
 STVQVDAEPGADAEFGADAEVEEQKQKQKRGFTGARKSARKLANQKRF
 SEPMPQVLSMQSSQRRRLHONGNV"
 1317..1322

PolyA_signal
 BASE COUNT 309 a 370 c 386 g 289 t
 ORIGIN

Query Match 59.4%; Score 554; DB 29; Length 1354;
 Best Local Similarity 89.7%; Pred. No. 0.00e+00;
 Matches 858; Conservative 0; Mismatches 58; Indels 41; Gaps 14;

Db 399 CTCTCAGCGTTCAGAGCTGTCCAGGAGTCCCTCAGGAGTCCGAGGAGAGTCCAC 458
 QY 1 CTCTCAGCGTTCAGAGCTGTCCAGGAGTCCCTCAGGAGTCCGAGGAGAGTCCAC 60
 Db 459 CCAGTCCAGACAGCTTCACTTAAGACGTGATCCCTCCAGTGAAGAGAGTCTTC 518
 QY 61 CCGGCTCGGGCAGCTTCACTTAAGACGTGATCCCTCCAGTGAAGAGAGTCTTC 120
 Db 519 CCGGCGCTCTCTGAGCTGATGACCAAGTATGAGCCAGTGGCTCCGCGGAGTCCAG 578
 QY 121 CCGGCGCTCTCTGAGCTGATGACCAAGTATGAGCCAGTGGCTCCGCGGAGTCCAG 180
 Db 579 CTGTCTGACCCAGAGCTTGTGAGCTGCTTTTACAGCCAAAGGCTCGGAGATGAG 638
 QY 181 CTGTCTGATCCAGAGCTCAGCTCCGGTGTCTTTACAGTGCAGAACTTCGGAGATCAG 240
 Db 639 CACCTGAAGAGAAATGCCGCTGTCCGAGCTTGTGCAAGTGCAGATCCAGAGCTGCG 698
 QY 241 CACCTGAAGAGAAATGCCGCTGTCCGAGCTTGTGCAAGTGCAGATCCAGAGCTGCG 300
 Db 699 GCAGATGCTGAGCTGGCGAGATGCTGAGAGCTGGCGAGATGAGAGAGTGGAG 758
 QY 301 GCAGATGCTGAGAGCTGTCGAGATG---AG---GC-C---GA---TGAG---TGAG 342
 Db 759 CAGAGAGCTGCGAGAAAGTGTGGGGCTTCACCGGGCCCGGAAATCAGCCCGGAG 818
 QY 343 CAGAGAGCTGCGAGAAAGTGTGGGGCTTCACCGGGCCCGGAAATCAGCCCGGAG 402
 Db 819 TTGGCCAAACAGAGGAGTTCAGTGTATGAGGAGTACTGCTGAGCATGAG 878
 QY 403 TTGGCCAAACAGAGGAGTTCAGTGTATGAGGAGTACTGCTGAGCATGAG 462
 Db 879 TCAAGTCAAGCGCGCGACCTGACCCAGCAAGTGTATGTGTAGCCAGAAAGAGCCCT 938
 QY 463 TCAAGTCAAGCGCGCGACCTGACCCAGCAAGTGTATGTGTAGCCAGAAAGAGCCCT 522
 Db 939 CCGAGCTGACCGGCGACCTGACCCAGCAAGTGTATGTGTAGCCAGAAAGAGCCCT 998
 QY 523 CCGAGCTGACCGGCGACCTGACCCAGCAAGTGTATGTGTAGCCAGAAAGAGCCCT 582
 Db 999 TCCACTCAAGAGCTGACCGGCGAGTCAAGTGTATGTGTAGCCAGAAAGAGCCCT 1058
 QY 583 TCCACTCAAGAGCTGACCGGCGAGTCAAGTGTATGTGTAGCCAGAAAGAGCCCT 638
 Db 1059 CCAGAGACCTGCTCCCGGCTGTGGCCCTCCCTCATGAGCATGTCCACAGATCT 1118
 QY 639 ---A-CACC---C-GC-----G--GCACCTCC-CACAGCATGTCCACCAATCT 680
 Db 1119 GTTGCTACATCAGAGTGTATTTTGTATTTCTCCAGTATCAATTTAATGCCCCATCT 1178
 QY 681 GTTGCTACATCAGAGTGTATTTTGTATTTCTCCAGTATCAATTTAATGCCCCATCT 740
 Db 1179 TCTCAGTATCTTTCGCCCTCATAGGGCAGTGAAGAGCAATGAATCAGACCTGG 1238
 QY 741 TCTTCTCATCTCTGCGCTCTCGTAGGGCAGTGAAGAGCAATGAATCAGACCTGG 800
 Db 1239 GGTTTGGCTCTCAGTCCCATTAAGTGTGTAAAGAGCTGTTTGTGAGTATG 1298
 QY 801 GGTTTGGCTCTCAGTCCCATTAAGTGTGTAAAGAGCTGTTTGTGAGTATG 860
 Db 1299 TTAAACAACTTCTCATTAAGTGTGTAAAGAGCTGTTTGTGAGTATG 1354
 QY 861 TTGGAACAACTTCTCATTAAGTGTGTAAAGAGCTGTTTGTGAGTATG 917

RESULT 5
 LOCUS MUSN23K 1047 bp mRNA
 DEFINITION Mouse N23K mRNA for N23K protein, complete cds. 14-MAR-1997

ACCESSION D50056
 NID 91857023
 KEYWORDS N23k protein; N23k.
 SOURCE Mus musculus neuroblastoma cell_line:NS20Y cDNA to mRNA.
 ORGANISM Mus musculus
 Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata;
 Vertebrate; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae;
 Murinae; Mus.
 REFERENCE 1 (bases 1 to 1047)
 AUTHORS Salto, Y., Maruyama, K., Salto, T.C. and Kawashima, S.
 TITLE N23k, a gene transiently up-regulated during neural
 differentiation, encodes a precursor protein for a newly identified
 neuropeptide nociceptin
 JOURNAL Biochem. Biophys. Res. Commun. 217 (2), 539-545 (1995)
 MEDLINE 96106851
 REFERENCE 2 (bases 1 to 1047)
 AUTHORS Salto, Y.
 TITLE Direct Submission
 JOURNAL Submitted (05-APR-1995) to the DDBJ/EMBL/GenBank databases. Yumiko
 Salto, Tokyo Metropolitan Institute for Medical Science, molecular
 biology; Honkomagome 3-18-22, Bunkyo-ku, Tokyo 113, Japan
 (E-mail: salto@rins.hoken.or.jp, Tel:03-3823-2101(ex.5248),
 Fax:03-5685-6609)
 COMMENT On Mar 1, 1997 this sequence version replaced gi:1217905.
 FEATURES
 source
 1..1047
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /cell_line="NS20Y"
 /cell_type="neuroblastoma"
 /gene="N23k"
 /note="N23k"
 /note="neuropeptide precursor"
 /codon_start=1
 /product="N23k protein"
 /db_xref="PID:d109395"
 /db_xref="PID:g1217905"
 /translation="MKILFCVLLSLSSVSPSPDCTCOEKLHPADSFNLRKTC
 ILCEKVEPRRLMTVCTKVMASGSGSPADPELYSALYOPKASEMHLKMPRV
 STVYRDAPGADAEFGADAEFGADAEVEOKOKLRGGTGAARKSKRLANOKRE
 SEFMROYLVLSWQSSORRRTLHONGNV"
 polyA_signal
 1040..1045
 BASE COUNT 237 a 315 c 263 g 232 t
 ORIGIN
 Query Match 59.0%; Score 550; DB 29; Length 1047;
 Best Local Similarity 89.6%; Pred. No. 0.00e+00;
 Matches 854; Conservative 0; Mismatches 58; Indels 41; Gaps 14;

301 GCAGATGCAGACCTGTGTCCAGATG---AG---GC-C---GA---TGAGG---TGAG 342
 456 CAGAAGCAGCTCAGAAAAAGTTTGGGGCTTACCGGGCCGGGAAATACACCCGGAAG 515
 343 CAGAAGCAGCTCAGAAAAAGTTTGGGGCTTACCGGGCCGGGAAATACACCCGGAAG 402
 516 TTGGCCACAGAGCGGTTCACTAGTATATAGAGCAGTACCTGGTCTGAGCATGAG 575
 403 TTGGCCACAGAGCGGTTCACTAGTATATAGAGCAGTACCTGGTCTGAGCATGAG 462
 576 TCAAGTCAACGGCGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 635
 463 TCAAGTCAACGGCGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 522
 636 CCAGCTGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 695
 523 CCAGCTGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 582
 696 TCCACTCCAGAGCTGCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 755
 583 TCCACACCAAGAGCTGCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 638
 756 CCAGCAGCAGCTCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 815
 639 ---A-CACC---C-GC-----G-GCAGCTGCC-CACAGATGCTCCACCAATCTCT 680
 816 GTTGCTACATCAGAGTATATTTTGTATTTCTCCAGCTACATTTTATGAGCCCATCT 875
 681 GTTGCTACATCAGAGTATATTTTGTATTTCTCCAGCTACATTTTATGAGCCCATCT 740
 876 TCTCACTCATCTTCTGCGGCTGATAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 935
 741 TCTTGTCTCATCTCTGCGGCTGATAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 800
 936 GTTGCTCATCTCTGCGGCTGATAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 995
 801 GTTGCTCATCTCTGCGGCTGATAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 860
 996 TTTAAACACACTTTCCTCACTTAACTTCTACTGAGGAA-TGGTTAATATAAAA 1047
 861 TTTGAAACACACTTTCCTCACTTAACTTCTACTGAGGAAATGTTAATATAAAA 913

RESULT 6
 LOCUS MPM3MENA 633 bp RNA ROD 10-SEP-1996
 DEFINITION M.musculus mRNA for prepronociceptin.
 ACCESSION X97373
 NID 91532039
 KEYWORDS prepronociceptin.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata;
 Vertebrate; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Mus.
 REFERENCE 1 (bases 1 to 633)
 AUTHORS Mollereau, C., Simons, M.J., Soularue, P., Lihers, F., Vassart, G.,
 Meunier, J.C. and Parmentier, M.
 TITLE Structure, tissue distribution, and chromosomal localization of the
 prepronociceptin gene
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 93 (16), 8666-8670 (1996)
 MEDLINE 96333281
 REFERENCE 2 (bases 1 to 633)
 AUTHORS Parmentier, M.
 TITLE Direct Submission
 JOURNAL Submitted (19-APR-1996) M. Parmentier, Universite Libre de
 Bruxelles, I R I B H N ULB Campus Erasme, 808 Route de Lennik, B-
 1070 Bruxelles, Belgium
 FEATURES
 source
 1..633
 /organism="Mus musculus"
 /strain="129/Sv"
 /db_xref="taxon:10090"


```

/clone.lib="lambda FIXII (Stratagene #946305)"
/clone="MG43"
<1..150
/number=2
join(<1..150,151..>633)
join(25..150,151..588)
/codon_start=1
/product="prepronocleptin"
/db_xref="PID:e24874"
/translation="MKILFCDVLLSLISVSSPCPRDLTQCEKILHPADSPNLKTC
ILQCEKVFPRPLMTCTKVMASGSGOLSPADPELVSAALYQKASSEMOKLRMPVR
SLVOYRDAEPGADAPGADAEVDAQOLQKRRFGFTGARKSARKLANOKRF
SEFMQYLVLSMOSQRRRLHONGNY"
sig_peptide
mat_peptide
25..69
70..585
/product="prepronocleptin"
/number=3
151..>633
exon
BASE COUNT      138 a      194 c      182 g      119 t
ORIGIN
Query Match      39.3%; Score 366; DB 29; Length 633;
Best Local Similarity 90.0%; Pred. No. 5.37e-201;
Matches 513; Conservative 0; Mismatches 39; Indels 18; Gaps 6;

Db 64 CTCTCAGCGGTTCAGAGCTGTCCAGAGCTCCTCAGCTGCCAGAGAAAGCTCCAC 123
    |||
QY 1 CTCTCAGCGGTTCAGAGCTGTCCAGAGCTCCTCAGCTGCCAGAGAAAGCTCCAC 60
    |||
Db 124 CCAGCTCCAGAGCTTCAACTTAAAGAGCTGCATCTCCAGTGTGAAGAGAGCTTTC 183
    |||
QY 61 CCGGCTCCGGGAGCTTCAAGCTGAGCTGATCTCCAGTGTGAAGAGAGCTTTC 120
    |||
Db 184 CCGGCTCCGGGAGCTTCAAGCTGAGCTGATCTCCAGTGTGAAGAGAGCTTTC 243
    |||
QY 121 CCGGCTCCGGGAGCTTCAAGCTGAGCTGATCTCCAGTGTGAAGAGAGCTTTC 180
    |||
Db 244 CTGCTGACCCAGAGCTTGTGAGCTGCTTACAGAGAAAGGCTCCGAGATGCAG 303
    |||
QY 181 CTGCTGATCCAGAGCTCAGCTCCGCTTACCAAGTCCGAGAGCTCCGAGATGCAG 240
    |||
Db 304 CACCTGAAGAGATGCCGCTGTCCGAGCTTGTGCAAGTGCAGAGTCCAGAGCTGC 363
    |||
QY 241 CACCTGAAGAGATGCCGCTGTCCGAGCTTGTGCAAGTGCAGAGTCCAGAGCTGC 300
    |||
Db 364 GCAGATGCTGAGCTGCGGAGATGCTGAGCTGCGGAGAGATGAGAGCTGAGAG 423
    |||
QY 301 GCAGATGCGAGAGCTTCCAGATG---AG---GC---GA---TGAGG---TGAG 342
    |||
Db 424 CAGAGACAGCTGCAGAAAGTTTGGGGCTTACCGGGGCGCGAAATCCAGCCCGAG 483
    |||
QY 343 CAGAAACAGCTGCAGAAAGTTTGGGGCTTACCTGGGGCGCGAAATCCAGCCCGAG 402
    |||
Db 484 TTGGCCACAGAGAGGTTCAAGTATGAGGCAAGTACCTGCTTGAAGATGCAG 543
    |||
QY 403 TTGGCCACAGAGAGGTTCAAGTATGAGGCAAGTACCTGCTTGAAGATGCAG 462
    |||
Db 544 TCAAGTCAAGCGCGGCGACCTGCAGCAGAAATGTAATGTAAGCCAGAGAGAGCCCT 603
    |||
QY 463 TCAAGTCAAGCGCGGCGACCTGCAGCAGAAATGTAATGTAAGCCAGAGAGAGCCCT 522
    |||
Db 604 CCCAGCTGCAGCGGCGACTGCAGCCATGA 633
    |||
QY 523 CCCAGCTGCAGCGGCGACTGCAGCCATGA 552
    |||

RESULT 7
LOCUS M25N27K 972 bp mRNA ROD 12-MAR-1997
DEFINITION Mouse N23K mRNA for developmental-regulated molecule, complete cds.
ACCESSION D50055
NID g1262587
KEYWORDS developmental-regulated molecule; N23K; neuropeptide.

```

```

SOURCE Mus musculus neuroblastoma cell_line: NS20Y cDNA to mRNA.
ORGANISM Mus musculus
REFERENCE Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
AUTHORS Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae;
Muriinae; Mus.
TITLE 1 (sites)
Saito, Y., Maruyama, K., Kawano, H., Hagiino-Yamagishi, K., Kawamura, K.,
Saito, T.C., and Kawashima, S.
JOURNAL Molecular cloning and characterization of a novel form of
MEDLINE neuropeptide gene as a developmentally regulated molecule
AUTHORS J. Biol. Chem. 271 (26), 15615-15622 (1996)
REFERENCE 96279082
AUTHORS 2. (bases 1 to 972)
Saito, Y.
JOURNAL Unpublished (1996)
REFERENCE 3 (bases 1 to 972)
AUTHORS Saito, Y.
JOURNAL Direct Submission
TITLE Submitted (05-APR-1995) to the DDBJ/EMBL/GenBank databases. Yumiko
Saito, Tokyo Metropolitan Institute for Medical Science, molecular
biology, Honkomagome 3-18-22, Bunkyo-ku, Tokyo 113, Japan
(E-mail: saito@etinsoken.or.jp, Tel: 03-3823-2101 (ex. 5248),
Fax: 03-5685-6609)
FEATURES
source
1..972
/organism="Mus musculus"
/db_xref="taxon:10090"
/cell_line="NS20Y"
/cell_type="neuroblastoma"
43..681
/gene="N23K"
43..681
/gene="N23K"
/codon_start=1
/product="developmental-regulated molecule"
/db_xref="PID:g1304162"
/translation="MKILFCDVLLSLISVSSPCPRDLTQCEKILHPADSPNLKTC
ILQCEKVFPRPLMTCTKVMASGSGOLSPADPELVSAALYQKASSEMOKLRMPVR
SLVOYRDAEPGADAPGADAEVDAQOLQKRRFGFTGARKSARKLANOKRF
SEFMQYLVLSMOSQRRRLHONGNYIYIPFACVHSKTRPVRIPSPRH"
BASE COUNT      213 a      291 c      246 g      222 t
ORIGIN
Query Match      38.1%; Score 355; DB 29; Length 972;
Best Local Similarity 83.5%; Pred. No. 4.23e-194;
Matches 773; Conservative 0; Mismatches 100; Indels 53; Gaps 27;

Db 82 CTCTCAGCGGTTCAGAGCTGTCCAGAGCTCCTCAGCTGCCAGAGAGCTCCAC 141
    |||
QY 1 CTCTCAGCGGTTCAGAGCTGTCCAGAGCTCCTCAGCTGCCAGAGAGCTCCAC 60
    |||
Db 142 CCAGCTCCAGAGAGCTTCAACTTAAAGAGCTGCATCTCCAGTGTGAAGAGAGCTTTC 201
    |||
QY 61 CCGGCTCCGGGAGAGCTTCAAGCTGAGAGCTGCATCTCCAGTGTGAAGAGAGCTTTC 120
    |||
Db 202 CCGGCTCCGGGAGAGCTTCAAGCTGAGAGCTGCATCTCCAGTGTGAAGAGAGCTTTC 261
    |||
QY 121 CCGGCTCCGGGAGAGCTTCAAGCTGAGAGCTGCATCTCCAGTGTGAAGAGAGCTTTC 180
    |||
Db 262 CCGGCTCCGGGAGAGCTTCAAGCTGAGAGCTGCATCTCCAGTGTGAAGAGAGCTTTC 321
    |||
QY 181 CCGGCTCCGGGAGAGCTTCAAGCTGAGAGCTGCATCTCCAGTGTGAAGAGAGCTTTC 240
    |||
Db 322 CACCTGAAGAGAGAGCTGCGGAGTGTCCGAGCTTGTGCAAGTGCAGATGCAGAGCTTGC 381
    |||
QY 241 CACCTGAAGAGAGAGCTGCGGAGTGTCCGAGCTTGTGCAAGTGCAGATGCAGAGCTTGC 300
    |||
Db 382 GCAGATGCTGAGAGCTGCGGAGTGTGAGAGCTGAGAGCTGAGAGCTGAGAGCTTGC 441
    |||
QY 301 GCAGATGCTGAGAGCTGCGGAGTGTGAGAGCTGAGAGCTGAGAGCTTGC 342
    |||

```


Db 442 CAGAGCAGCTGCGAAGAAAGGTTGGGGGCTTACCGGGGCGCCGAATCAGCCCGAG 501
|||||
Oy 343 CAGAGCAGCTGCGAAGAAAGGTTGGGGGCTTACCGGGGCGCCGAATCAGCCCGAG 402
|||||
Ub 502 TTGGCCACAGCAGCAGCTGAGTGTATGAGGAGTACTGCTGCTGAGCATGCG 561
|||||
Oy 403 TTGGCCACAGCAGCAGCTGAGTGTATGAGGAGTACTGCTGCTGAGCATGCG 462
|||||
Db 562 TCAAGTCAAGCCGGCGCGCCCTGACCAAGATGCA-----T--CCAGGTATCCCCG 614
|||||
Oy 463 TCAAGTCAAGCCGGCGCGCCCTGACCAAGATGCA-----T--CCAGGTATCCCCG 522
|||||
Db 615 TACAGCATGT--GTTCATCTCAAGACCTG--CAGGCCG--GAGTCAGATTC--C--TCCC 666
|||||
Oy 523 CCCAGCTGACCCGGCGCGCCCTGACCAAGATGCA-----T--CCAGGTATCCCCG 582
|||||
Db 667 TCC-C-CGAGGAGCTG--AGCAGCC-A--CAGCA--CCTCCCTGCCG--GT--TGTG-C 713
|||||
Oy 583 TCCAGACCAACAGCTGCGAGCCGGGAGTCAAGATTCCTCTCCCTGAGCAGCATGAC 642
|||||
Db 714 CC-CT-C-CCTC--ATGCGATGTCTCAGCAGATCTGTGTACATCAGAGTATTT 767
|||||
Oy 643 CCGGCGGAGCTCCCGCAGCATGTCTCAGCAGATCTGTGTACATCAGAGTATTT 702
|||||
Db 768 TTGTATTTCTCCAGCTAATATTTATGCGCCCATCTCTCAGCATCTCTGCTCT 827
|||||
Oy 703 TTGTATTTCTCCAGCTAATATTTATGCGCCCATCTCTCTCAGCATCTCTGCTCT 762
|||||
Db 828 CATAGGCGCAGTGAAGAGATGAATCAGACCTGGGTTTGGCTTCCCTCAGCGCAT 887
|||||
Oy 763 CAGAGGCGCAGTGAAGAGATGAATCAGACCTGGGTTTGGCTTCCCTCAGCGCAT 822
|||||
Db 888 AACTGTTTGTAAAGAGCTGTCTTTTGAATGATCTGTAAACAATTCTTCATTA 947
|||||
Oy 823 AACTGTTTGTAAAGAGCTGTCTTTTGAATGATCTGTAAACAATTCTTCATTA 882
|||||
Db 948 AACTTCTACTAGCAAA-TGCTTAT 972
|||||
Oy 883 AACTTCTACTAGCAAAATGCTTAT 908
|||||

RESULT 8
LOCUS HSPHNRNA 1015 bp RNA PRI 10-SEP-1996
DEFINITION H.sapiens mRNA for prepronocleptin.
ACCESSION X97370
NID g1531988
KEYWORDS prepronocleptin.
SOURCE human.
ORGANISM Homo sapiens
Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Euteleostomi; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 1015)
Moller, J.C., Simons, M.J., Souleau, P., Liners, F., Vassart, G.,
Moller, J.C., and Parmentier, M.,
Structure, tissue distribution, and chromosomal localization of the
prepronocleptin gene
Proc. Natl. Acad. Sci. U.S.A. 93 (16), 8666-8670 (1996)
JOURNAL 2 (bases 1 to 1015)
MEDLINE 96323281
REFERENCE Direct Submission
AUTHORS Submitted (19-APR-1996) M. Parmentier, Universite Libre de
JOURNAL Bruxelles, I R I B H N ULB Campus Erasme, 808 Route de Lennik, B-
1070 Bruxelles, BELGIUM
FEATURES
SOURCE
1 1015
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="lambda DASHII (Stratagene #94520)"
/clone="HG91"
/clone="HGEX107"
/chromosome="8"
/map="p21"

/map="1.9 cr from WI-1172 marker"
<1..149
/number=2
mRNA join(<1..149,150..601,602..>1015)
sig_peptide 24..68
CDS join(24..149,150..554)
/codon_start=1
/product="prepronocleptin"
/db_xref="PID:e24477"
/db_xref="PID:g1531988"
/translation="MKVLCIDLILSLFSSVSSCORDLTCOEKTLHPALDSFLEVC
SLREQEPEPEPEHEHENGEMQKOLKRFGEFTGARKSARKLNQKRFSEPMROYLVLS
MQSSQRRRLHQNV"
mat_peptide 69..551
/product="prepronocleptin"
exon 150..601
/number=3
exon 602..>1015
/number=4
polya_site 1015
BASE COUNT 212 a 304 c 264 g 235 t
ORIGIN
Query Match 36.1%; Score 336; DB 26; length 1015;
Best Local Similarity 82.5%; Pred. No. 3,33e-182;
Matches 534; Conservative 0; Mismatches 96; Indels 17; Gaps 9;
Db 64 TCTCCAGTGTGTGAGAGAGTGTGAGAGGAGGAGCTCTCAATCAGAGAGAGAGAGTCCAGC 123
|||||
Oy 2 TCTCCAGTGTGTGAGAGAGTGTGAGAGGAGGAGCTCTCAATCAGAGAGAGAGTCCAGC 61
|||||
Db 124 CAGCCTGAGACAGCTTGACCTGAGAGTGTGACATCTCGAGTGTGAGAGAGAGTCTTCC 183
|||||
Oy 62 CAGCCTGAGAGAGCTTGACCTGAGAGTGTGACATCTCGAGTGTGAGAGAGAGTCTTCC 121
|||||
Db 184 CAGACCCCTCTGAGCTGCATGACACCAAGTGCATGAGAGAGTCTTGGCAGCTCAGCC 243
|||||
Oy 122 CCCCCTCTCTGAGCTGTGACACCAAGGAGGAGGAGCTGCATGAGAGAGCTCAGCC 181
|||||
Db 244 CTGCGGCCCGCAGAGCATGTGGGGGCTGCTCTCTCAACGCGAGAGCTTGGAGATGAGC 303
|||||
Oy 182 CTGCTGATTCAGAGCTCAGCTCAGCTCTCTTTCAGAGTGAAGAGCTCGAGATGAGC 241
|||||
Db 304 ATCTGCGCGAGATGCGCCGAGATGCGGAGCTGTTCCAGAGAG--AG--GAAGAGCCGAGC 360
|||||
Oy 242 ACCGTGAAGAGATGCCCGCTGTCAGAGTGTGTCAGAGCCCGAGAGAGAGCTGAGG 301
|||||
Db 361 CTG--GCA----TG--G-AG--GAGGCTGTGATGAGAGCAGACAGCTGCAGAGA 408
|||||
Oy 302 CAGATGAGAGAGCTGTGAGAGATGAGAGCCGATGAGTGAAGAGAGAGCTGCAGAGAA 361
|||||
Db 409 GATTGGGGGCTTACCGGGGCGCCGGAAGTGGCCAGAGAGTGGCCAAATCAGAGAGGGGT 468
|||||
Oy 362 GATTGGGGGCTTACCGGGGCGCCGGAAGTGAAGCCGGAATGGCCAAACAGAGAGGGGT 421
|||||
Db 469 TCACTGAGTTATGAGCAATACTGTGCTGAGCATGAGATCCAGCAGAGCGCGGCGCA 528
|||||
Oy 422 TCACTGAGTTATGAGCAGTACTGCTGAGCATGAGATCCAGCAGAGCGCGGCGCA 481
|||||
Db 529 CCGTGCACAGAGATGTAATGTGAGCCGGAAGGGGGCTCTCCAGCTGTAGCGGCA 588
|||||
Oy 482 CTCTGACAGAGATGTAATGTGAGCCAGAGAGAGC--C--CCTCCAGCTGACCGGCA 539
|||||
Db 589 CTGGAACCAAGAGCGGTCAGAGTATCCCGCAACAGCATGTGTGACCCCAACACTGC 648
|||||
Oy 540 CTGGAACCAAGAGCATCCAGAGTATCCCGCAACAGCATGTGTGACCCCAACACTGC 599
|||||
Db 649 CGCCTGGGAATCAGATCTCTTCCCAAGAGCAGTGAAGCGCTGC 695
|||||
Oy 600 AGGCGGGAGTCAAGATTCCTCTTCCCTGAGGAGCATGAGACACCGCG 646
|||||

[illegible][illegible]

Db	660	GCACCAAGATGATGGCCACGTGGCTCCGGGACACTCAGCCCTGTCAGACCAAGCTTTGT	719
Oy	143	GCACCAAAAGCCATGCGCCAGTGACTCTGAGACACTCAGCCCTGCTGATCCAGAGCTCAGCT	202
Db	720	CAGCTGCTCTTTTACCAGCCAAAGGCGCTCGGAGATGCAGCACCTGAAAGAAATGCGCGTG	779
Oy	203	CCGGCTGCTTTTACCAGTGCAGAAAGCGCTCGGAGATGCAGACCTGAAAGAAATGCGCGTG	262
Db	780	TCCGAGCTTGTGCAATGCGAGATGCAGACCTCGCGGCAGATCTGAGCTTGGCGCAG	839
Oy	263	TCAGAGAGTGTGTGCAGAACCCGAGAGCGAGACCTCTGAGGCAATGCAGAGCTGTTCGAG	322
Db	840	ATGTGAGCCCTGGCCACAGTGAACGCTGAGSAGTGGAGAGAGAGCAGTGCAGAAAAGT	899
Oy	323	ATG---AG-----GC-C-----GA---TGAG---TGAGAGAGAGAGCTGCAGAAAAGT	364
Db	900	TTGGGGGCTTCAACCGGGGCGCGGAAATCATGCCCGGAAGTTGGCCAAACCAAGAGCGTTCA	959
Oy	365	TTGGGGGCTTCACTGTGGGCGCCGGAAGTCAAGCCCGGAAGTTGGCCAAACCAAGAGCGGTTC	424
Db	960	GTGAGTTTATGAGGAGTACCTGAGTCTGAGAGTCAAGTCAAGTCAAGCGCGGCGACACC	1019
Oy	425	GTGAGTTTATGAGGAGTACCTGAGTCTGAGAGTCAAGTCAAGTCAAGCGCGGCGACACC	484
Db	1020	TGCACAGAAATGCTAATGTGTAGCCAGAGAGAGCCCTCCACGTGCACACGGCCACTGCA	1079
Oy	485	TGCACAGAAATGCTAATGTGTAGCCAGAGAGAGCCCTCCACGTGCACACGGCCACTGCA	544
Db	1080	ACCCATGAG 1088	
Oy	545	ACCCATGAG 553	

RESULT	11	AB005251	1209 bp	mRNA	MAM	09-APR-1998
LOCUS		Bos taurus mRNA for nociceptin/orphanin FQ precursor, complete cds.				
DEFINITION		AB005251				
ACCESSION		93041768				
NID		nociceptin/orphanin FQ precursor.				
KEYWORDS		Bos taurus cDNA to mRNA.				
SOURCE		Bos taurus				
ORGANISM		Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Artiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.				
REFERENCE		1 (sites)				
AUTHORS		Okuda-Ashtaka,E., Minami,T., Tachibana,S., Yoshinara,Y., Nishitani,Y., Kimura,T. and Ito,S.				
TITLE		Nocistatin, a peptide that blocks nociceptin action in pain transmission				
JOURNAL		Nature 392 (6673), 286-289 (1998)				
MEDLINE		98180619				
REFERENCE		2 (bases 1 to 1209)				
AUTHORS		Okuda-Ashtaka,E.				
TITLE		Direct Submission				
JOURNAL		Submitted (26-JUN-1997) to the DDBJ/EMBL/GenBank databases. Emiko Okuda-Ashtaka, Kansai Medical University, Department of Medical Chemistry, 10-15 Fumizono, Moriyuchi, Osaka, JAPAN, Moriyuchi, Osaka 570, Japan (E-mail:ashtakatakai.kmu.ac.jp, Tel:06-9992-1001, Fax:06-997-1781)				
FEATURES		Location/Qualifiers				
Source		1..1209				
		/Organism="Bos taurus"				
		/db_xref="taxon:9913"				
		210..740				
CDS		/codon_start=1				
		/product="nociceptin/orphanin FQ precursor"				
		/db_xref="PID:d102634"				
		/db_xref="PID:g3041769"				
		/translation="MKILFCDLLILFSSVSSGCGDCLVCREKRLPTLDSLEMKQ ILCEERKAFSLPTPTCKVARGSWGLSPADPDVVAALQOPRASEMQLKRPDRVY SLFQRKTRPGLFLEVGVEIDKQLOKRRGCTGAKRKSRLKLANQRSEFNRQIVLVS MOSSQRRTTLHONGNA"				
BASE COUNT		240 a 359 c 341 g 269 t				

[illegible]

THIS PAGE BLANK (USPTO)

-1- SIMILARITY: BELONGS TO THE OPIOIDS NEUROPEPTIDES PRECURSORS
 CC FAMILY.
 CC EMBL: U48262; G1185012; -.
 DR EMBL: X97375; E244795; -.
 DR EMBL: S79730; G1172239; -.
 DR PROSITE: PS01252; OPIOIDS_PRECURSOR; 1.
 KW OPIOID PEPTIDE; NEUROTRANSMITTER; NEUROPEPTIDE;
 KM CLEAVAGE ON PAIR OF BASIC RESIDUES; SIGNAL; REPEAT.
 FT SIGNAL 1
 FT PEPTIDE 98 132 POTENTIAL.
 FT PEPTIDE 135 151 NEUROPEPTIDE 1 (PROBABLE).
 FT PEPTIDE 134 170 NOCICEPTIN (ORPHANIN FQ).
 FT DOMAIN 109 120 NEUROPEPTIDE 2 (PROBABLE).
 FT REPEAT 109 114 2 x 6 AA TANDEM REPEATS OF D-A-E-P-X-A.
 FT REPEAT 115 120 1.
 SQ SEQUENCE 181 AA; 20478 MW; 1640366F CRC32;
 Query Match 42.4%; Score 1978; DB 1; Length 181;
 Best Local Similarity 55.1%; Pred. No. 0.00e+00;
 Matches 277; Conservative 133; Mismatches 93; Indels 0; Gaps 0;
 DB L S S V F S S C P E D C L T C Q E R L H
 DC 40 YNNMSWNGTITTTNMSNTGYCCNGARAYTGTYTNACTGACARARMGNTNAY 99
 QY 1 CCTCCAGGCTGTCAGCAGCTGCCAGAGCTGCTCCTGCTGCTGCTGCTGCTGCT 60
 QT L S S V F S S C P E D C L T C Q E R L H
 DB P A P G S F N L K L C I L O C E E K V F
 DC 100 CGNGCCGNGWNTTAYTNTGATNTGATNTGATNTGATNTGATNTGATNTGATNT 159
 QY 61 CGGCTCCGGGAGCTGACCTGACCTGACCTGACCTGACCTGACCTGACCTGACCT 120
 QT P A P G S F N L K L C I L O C E E K V F
 DB P R P L W T L C T K A M A S D S E Q L S
 DC 160 CGNGCCGNGWNTTAYTNTGATNTGATNTGATNTGATNTGATNTGATNTGATNT 219
 QY 121 CCCCCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
 QT P R P L W T L C T K A M A S D S E Q L S
 DB P A D P E L T S A A L Y O S K A S E M Q
 DC 220 CGNGCCGNGWNTTAYTNTGATNTGATNTGATNTGATNTGATNTGATNTGATNT 279
 QY 181 CCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
 QT P A D P E L T S A A L Y O S K A S E M Q
 DB H L K R M P R V R S V V Q A R D A E P E
 DC 280 CAYTNARMGNTGATNTGATNTGATNTGATNTGATNTGATNTGATNTGATNT 339
 QY 241 CACCTGAAGAAATGCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
 QT H L K R M P R V R S V V Q A R D A E P E
 DB A D A E P V A D E A D E V E Q K O L Q K
 DC 340 CGNGCCGNGWNTTAYTNTGATNTGATNTGATNTGATNTGATNTGATNTGATNT 399
 QY 301 GCACTGACAGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
 QT A D A E P V A D E A D E V E Q K O L Q K
 DB R F G G F T G A R K S A R K L A N O K R
 DC 400 MGNITGNGGNTTAYTNTGATNTGATNTGATNTGATNTGATNTGATNTGATNT 459
 QY 361 AGGTTGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
 QT R F G G F T G A R K S A R K L A N O K R
 DB F S E F M R Q Y L V L S M Q S S Q R R R
 DC 460 TTYWSNGARTYATGMCNARTYATGNTYTNMSNATGABMSNATGABMSNATGAB 519
 QY 421 TTCAGTGAGTTATGAGGAGTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
 QT F S E F M R Q Y L V L S M Q S S Q R R R
 DB T L H Q N G N
 DC 520 ACNNTNCACTCARAYGNGNAYGT 542
 QY 481 ACTCTGACACCAAGATGTATGT 503

QT T L H Q N G N
 RESULT 2
 AC PNOG_MOUSE STANDARD; PRT; 187 AA.
 DT 064387; Q61105; Q61938;
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE NOCICEPTIN PRECURSOR (ORPHANIN FQ) (PNOG) (N23K / N27K).
 GN PNOG OR NPNCI.
 OS MUS MUSCULUS (MOUSE).
 CC EDUAROTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA.
 CC EUTHERIA; RODENTIA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN.
 RX MEDLINE; 96216718.
 RA HOOTANI T., NISHI M., TAKESHIMA H., NUKADA T., SUGIMOTO T.;
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 219:714-719(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 96106851.
 RA SAITO Y., MARYAMA K., SAITO T.C., KAWASHIMA S.;
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 217:539-545(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-129/SV.
 RX MEDLINE; 96323281.
 RA MOLLEREAU C., SIMONS M.-J., SOULARNE P., LINERS F., VASSART G.,
 RL MONTIER J.-C., PARMENTIER M.;
 RN PNOG. NATL. ACAD. SCI. U.S.A. 93:8666-8670(1996).
 RN [4]
 RP SEQUENCE FROM N.A. (LONG FORM).
 RX MEDLINE; 96279082.
 RA SAITO Y., MARYAMA K., KAWANO H., HAGINO-YAMAGISHI K.,
 RA KANAKURA K., SAITO T.C., KAWASHIMA S.;
 RL J. BIOL. CHEM. 271:15615-15622(1996).
 RN [5]
 RP SEQUENCE OF 79-187 FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE; 96207555.
 RA PAN Y.-X., XU J.-J., PASTERNAK G.W.;
 RL BIOCHEM. J. 315:11-13(1996).
 CC -1- FUNCTION: NOCICEPTIN IS THE LIGAND OF THE OPIOID RECEPTOR-LIKE
 RECEPTOR (OPR1). IT MAY ACT AS A TRANSMITTER IN THE BRAIN BY
 MODULATING NOCICEPTIVE AND LOCOMOTOR BEHAVIOR. MAY BE INVOLVED
 IN NEURONAL DIFFERENTIATION AND DEVELOPMENT. WHEN ADMINISTERED
 INTRACEREBROVENTRICULARLY, NOCICEPTIN INDUCES HYPERALGESIA AND
 DECREASES LOCOMOTOR ACTIVITY.
 CC -1- SUBCELLULAR LOCATION: SECRETED INTO THE ENDOPLASMIC RETICULUM.
 CC -1- TISSUE SPECIFICITY: BRAIN AND SPINAL CORD. LOW LEVELS IN
 KIDNEY AND SPLEEN.
 CC -1- DEVELOPMENTAL STAGE: IN EMBRYONIC BRAIN, FIRST DETECTED AT DAY 14
 AND IN POSTNATAL BRAIN, LEVELS INCREASE IN DAY 1 AND DAY 18.
 CC -1- LEVELS DECREASE SIGNIFICANTLY IN ADULTS.
 CC -1- PTM: SPECIFIC ENZYMATIC CLEAVAGES AT PAIRED BASIC RESIDUES
 PROBABLY YIELD OTHER ACTIVE PEPTIDES BESIDES NOCICEPTIN.
 CC -1- PTM: THE N-TERMINAL CONTAINS 6 CONSERVED CYSTEINES THOUGHT TO
 BE INVOLVED IN DISULFIDE BONDING AND/OR PROCESSING.
 CC -1- ALTERNATIVE PRODUCTS: A LONGER FORM, N7K IS PROBABLY PRODUCED
 BY ALTERNATIVE SPLICING.
 CC -1- SIMILARITY: BELONGS TO THE OPIOIDS NEUROPEPTIDES PRECURSORS
 FAMILY.
 CC EMBL: D82866; D1012281; -.
 DR EMBL: D50056; D1009395; -.
 DR EMBL: X97373; E244874; -.
 DR EMBL: X97371; E244782; -.
 DR EMBL: X97372; E244782; JOINED.
 DR EMBL: U44027; G1335870; -.
 DR EMBL: D50055; G1304162; -.
 DR MGD; MGI:105308; NPNCI.
 DR PROSITE; PS01252; OPIOIDS_PRECURSOR; 1.

Query March	Score 170;	DB 1;	Length 407;
Best Local Similarity	43.9%;	Pred. No. 6.54e-05;	
SEQUENCE	407 AA;	46617 MM;	1EFCAGCF CRC32;
DOMAIN	60	241	GLU-RICH (ACIDIC).
NET			

	MATCHES	98;	Conservative	25;	Mismatches	47;	Indels	2;	Gaps	2
Dd	E E A E E E A E E E A E E A E E E									
Dt	RGARGARCCNCGARGARGARGCNGARGARGAR-GCNGARGARGCNCARGARG	448								
Oy	: : : : : : : : : : : : : : :									
Ot	AGATGCACCTGATGAAGATGCCGCTGCAGAGTGTGGTCAGACCCGACAGCAG	292								
Dd	A A A P E E N A A C Q E V Y Q A R D A E									
Dt	ARGCNGARGARGCNCARGARGCNCARGARGCNCARGARGCNCARGARGCNC	508								
Oy	: : : : : : : : : : : : : : :									
Ot	AGCCTGAGGCGAATGACAGCTGCGCATGAGAGCCGATGAGAGTGAG-CAGAAGC	351								
Dd	A E E									
Dt	509 ARGCGARGARG 520									
Oy	: : :									
Ot	352 CTGCGAAGAGG 363 A E K									
RESULT	10									
ID	RVRN.RABIT STANDARD; PRF: 5037 AA.									
AC	P11716:									
Dt	01-OCT-1989 (REL. 12, CREATED)									
Dt	01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)									
Dt	15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)									
De	RYANODINE RECEPTOR, SKELETAL MUSCLE (SKELETAL CALCIUM RELEASE CHANNEL).									
Cn	RVRL									
Os	ORYZOLAGUS CUNICULUS (RABBIT).									
Sc	EUAROTIA: METALCOA: CHORDATA: VERTEBRATA: TETRAPODA: MAMMALIA:									
Oc	EUTHERIA: LAGOMORPHA.									
Nr	[1]									
Rp	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.									
Rc	TISSUE-SKELETAL MUSCLE;									
Rx	MEDLINE: 89262082.									
Ra	TAKESHIMA H., NISHIMURA S., MATSUMOTO T., ISHIDO H., KANGAWA K.,									
Ra	MINAMINO N., MASUO H., UEDA M., HANAOKA M., HIROSE T., NUMA S.;									
Rl	NATURE 339:439-445(1989).									
Rn	[2]									
Rp	SEQUENCE FROM N.A.									
Rc	TISSUE-SKELETAL MUSCLE;									
Rx	MEDLINE: 90130482.									
Ra	ZOZZARO F., FUJII J., OTSU K., PHILLIPS M., GREEN N.M., LAI F.-A.,									
Ra	MEISSNER G., MACLENNAN D.H.;									
Rl	J. BIOL. CHEM. 265:2244-2256(1990).									
Rn	[3]									
Rp	PHOSPHORYLATION OF SER-2843.									
Rc	MEDLINE: 93120178.									
Ra	SUGO Y., MAUDER-FOGY I., PLANK B., BERTEL O., WYSKOVSKY W.,									
Ra	HOHENEGGER M., HELLMANN G.;									
Rl	BIOCHIM. BIOPHYS. ACTA 1175:193-206(1993).									
Cc	-1- FUNCTION: COMMUNICATION BETWEEN TRANSVERSE-TUBULES AND									
Cc	SARCOPASMIC RETICULUM; CONTRACTION OF SKELETAL MUSCLE IS									
Cc	TRIGGERRED BY RELEASE OF CA++ FROM SR FOLLOWING DEPOLARIZATION OF									
Cc	T-TUBULES.									
Cc	-1- SUBUNIT: HOMOTETRAMER (POTENTIAL).									
Cc	-1- TISSUE SPECIFICITY: FAST- OR SLOW-TWITCH SKELETAL MUSCLE.									
Cc	-1- THE CALCIUM RELEASE CHANNEL IS MODULATED BY CA++, MG++, ATP, AND									
Cc	CALMODULIN.									
Cc	-1- THE CALCIUM RELEASE CHANNEL RESIDES IN THE C-TERMINAL									
Cc	REGION WHILE THE REMAINING PART OF THE PROTEIN CONSTITUTES THE									
Cc	'FOOT' STRUCTURE SPANNING THE JUNCTIONAL GAP BETWEEN THE SR AND									
Cc	THE T-TUBULE. IT IS POSSIBLE THAT THE FOOT STRUCTURE INTERACTS									
Cc	WITH THE CYTOPLASMIC REGION OF THE DIHYDROPYRIDINE RECEPTOR.									
Cc	-1- RANDONINE IS AN ALKALOID THAT BINDS TO THE CA-RELEASE CHANNEL IN									
Cc	JUNCTIONAL SR AND MODULATES ITS ACTIVITY.									

KM	RECEPOTRANSMEMBRANE IONIC CHANNEL; CALCIUM CHANNEL; REPEAT;
KW	PHOSPHORYLATION; GLYCOPROTEIN.
FT	MOD_RES ?1 ?1 BLOCKED.
FT	DOMAIN 1 3123 CYTOPLASMIC.
FT	TRANSMEM 3124 3144 M' (POTENTIAL).
FT	TRANSMEM 3188 3206 M'' (POTENTIAL).
FT	TRANSMEM 3985 4004 M1 (POTENTIAL).
FT	TRANSMEM 4023 4041 M2 (POTENTIAL).
FT	TRANSMEM 4277 4300 M3 (POTENTIAL).
FT	TRANSMEM 4342 4362 M4 (POTENTIAL).
FT	TRANSMEM 4559 4580 M5 (POTENTIAL).
FT	TRANSMEM 4648 4671 M6 (POTENTIAL).
FT	TRANSMEM 4789 4820 M7 (POTENTIAL).
FT	TRANSMEM 4837 4856 M8 (POTENTIAL).
FT	TRANSMEM 4879 4898 M9 (POTENTIAL).
FT	TRANSMEM 4914 4937 M10 (POTENTIAL).
FT	DOMAIN 1873 1913 GLU-RICH (ACIDIC).
FT	DOMAIN 842 2959 6 X APPROXIMATE REPEATS.
FT	REPEAT 842 955 1.
FT	REPEAT 956 1069 2.
FT	REPEAT 1345 1360 3 (INCOMPLETE).
FT	REPEAT 1373 1388 4 (INCOMPLETE).
FT	REPEAT 2726 2845 5.
FT	REPEAT 2846 2959 6.
FT	MOD_RES 2843 2843 PHOSPHORYLATION (BY CAK AND CGPK).
FT	MOD_RES 3952 3952 PHOSPHORYLATION (POTENTIAL).
FT	MOD_RES 4323 4323 PHOSPHORYLATION (POTENTIAL).
FT	SIMILAR 4628 4861 WITH ACHR-SUBUNITS TRANSMEMBRANE SEGMENTS M2 AND M3.
FT	CARBOHYD 3466 3466 POTENTIAL.
FT	CARBOHYD 3909 3909 POTENTIAL.
FT	CARBOHYD 3950 3950 POTENTIAL.
FT	CARBOHYD 4149 4149 POTENTIAL.
FT	CARBOHYD 4864 4864 POTENTIAL.
FT	CONFLICT 2015 E -> D (IN REF. 2).
FT	CONFLICT 3481 3485 MISSING (IN REF. 2).
SO	SEQUENCE 5037 AA: 565238 MW: 545466 CRC32;
Query Match 3.6%; Score 166; DB 1; Length 5037;	
Best Local Similarity 43.2%; Pred.No.2,29e-04;	
Matches 41; Conservative 16; Mismatches 38; Indels 0; Gaps 0	
Db	L K R K L G V D G E E E L V P E P E P
Dt	13416 HYTNAAARGNARFNTNGTNGYAGGANGARBARARTNTGCCNGARCNGARCC 13475
Oy	243 CCTGAAGGATGGCGCGCTGTCAAGAGTGGTGCAGAACCCGAGCGCAGCGCTGAGGC 302
Qt	L K R M P R V R S V V Q A R D A E P E A
Db	E P E P E P E K A D E
Dt	13476 NGARCCNGARCCNGARCCNGARARAGGANGAAGANG 13510
Oy	303 AGATCGAGAGCGCTGTGCGCATGAGGAGCGCGATGAGG 337
Qt	D A E P V A D E A D E
RESULT 11	
ID	NDBB.HUMAN STANDARD; PRF: 254 AA.
AC	P01213;
Dt	21-JUL-1986 (REL. 01, CREATED)
Dt	21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
Dt	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE	BETA-MEIODOPHIN-DYNORPHIN PRECURSOR (PROENKEPHALIN B PRECURSOR (PREPRODYNORPHIN)).
GN	PDYN
OS	HOMO SAPIENS (HUMAN).
OC	EURAROTIA METAFOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC	EUTHERIA; PRIMATES.
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE: 84068211.
RA	HOIKIRAMA S., TAKAI T., TOTOSATO M., TAKAHASHI H., NOYA M.,
RA	KAKIDANI H., KUBO T., HIROSE T., INAYAMA S., HAYASHIDA H., MIYATA T.,
RA	NODA S.;

Query Match	3.5%;	Score 165;	DB 1;	Length 254;
Best Local Similarity	38.9%;	Pred. No. 3.12e-04;		
Matches	49;	Conservative	20;	Mismatches 57;
			Indels	0;
			Gaps	0.

ID	RESULT	12	STANDARD:	PRT:	420 AA.
P2CG.MOUSE					
061074:					
15-JUL-1998	(REL. 36, CREATED)				
15-JUL-1998	(REL. 36, LAST SEQUENCE UPDATE)				
15-JUL-1998	(REL. 36, LAST ANNOTATION UPDATE)				
PROTEIN PHOSPHATASE 2C GAMMA ISOPFORM (EC 3.1.1.3, 16)	(PP2C-GAMMA)				
(FIBROBLAST GROWTH FACTOR INDUCIBLE PROTEIN 13) (F1N13)	(FRAGMENT).				
PM1C OR PIN13.					
MUS MUSCULUS (MOUSE).					
EUKARYOTA: METAZOA: CHORDATA: VERTEBRATA: TETRAPODA: MAMMALIA;					
EUTHERIA: MENTAZOA.					
[1]					
SEQUENCE FROM N.A.					
MEDLINE: 96226358.					
GUTHRIE M.A., SELDIN M., BASILICO C.;					
ONCOGENE 12:1267-1278(1996).					
-1 CATALYTIC ACTIVITY: A PHOSPHOPROTEIN + H(2O) -> A PROTEIN + ORTHOPHOSPHATE (THIS ENZYME IS SERINE/THREONINE SPECIFIC).					
-1 COPACITOR: BINDS TWO MAGNESIUM OR MANGANESE IONS (BY SIMILARITY).					
-1 SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).					
-1 TISSUE SPECIFICITY: HIGHLY EXPRESSED IN TESTIS. LOW LEVEL OF					

Query Match	3.5%	Score 161;	DB 1;	Length 420;
Best Local Similarity	39.1%;	Pred. No. 1.07e-03;		
Matches	52;	Conservative	31;	Mismatches 48;
			Indels	2;
			Gaps	2;

D _b	D	D	D	E	
D _t	544	GA	GA	GA	GA
O _Y	351	GCT	GCA	GAA	AGG
O _t		A	A	E	K

RESULT		13		
ID	P2CG_BOVIN	STANDARD;	PRT;	543 AA.
AC	P79126;			
DT	15-JUL-1998	(REL. 36, CREATED)		
DI	15-JUL-1998	(REL. 36, LAST SEQUENCE UPDATE)		
DT	15-JUL-1998	(REL. 36, LAST ANNOTATION UPDATE)		
DE	PROTEIN PHOSPHATASE 2C GAMMA ISOFORM (EC 3.1.3.16) (PP2C-GAMMA)			
DE	(MAGNESIUM-DEPENDENT CALCIUM INHIBITABLE PHOSPHATE) (MCPP).			
GN	PMIC.			
OS	BOS TAURUS (BOVINE).			
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;			
CC	EUETHERIA; ARTIODACTYLIA.			
RN	[1]			
R2	SEQUENCE FROM N.A.			
RA	HUANG C.Y., QIN R.;			
RL	SUBMITTED (DEC-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.			
CC	-1- CATALYTIC ACTIVITY: A PHOSPHOPROTEIN + H(2)O = A PROTEIN +			
CC	-1- ORGANOPHOSPHATE (THIS ENZYME IS SERINE/THREONINE SPECIFIC).			
CC	-1- COFACTOR: BINDS TWO MAGNESIUM OR MANGANESE IONS (BY SIMILARITY).			
CC	-1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).			
DR	EMBL: U01159; GI754694; -.			
PR	PROSITE: PS01032; PP2C; 1.			
KW	HYDROLASE; MAGNESIUM; MANGANESE; MULTIGENE FAMILY.			
FT	METAL	40	40	
FT	METAL	41	41	MANGANESE 1 (BY SIMILARITY).
FT	METAL	41	41	MANGANESE 1 (BY SIMILARITY).
FT	METAL	60	60	MANGANESE 1 AND 2 (BY SIMILARITY).
FT	METAL	439	439	MANGANESE 2 (BY SIMILARITY).
FT	METAL	494	494	MANGANESE 2 (BY SIMILARITY).
FT	DOMAIN	258	324	MANGANESE 2 (BY SIMILARITY).
FT	DOMAIN	266	272	ASP/GLU-RICH ACIDIC.
FT	DOMAIN	306	309	POLY-GLU.
Q	SEQUENCE	543 AA:	58708 MM:	RRORNOEF CRCT?.

Query Match 3.5%; Score 161; DB 1; Length 543;
Best Local Similarity 39.8%; Pred. No. 1.07e-03;
Matches 55; Conservative 30; Mismatches 48; Indels 2; Gaps 2;

DB D E A E E E E D S E E C S E E E D G
Dc 791 SNGAYGARGCNGARGARGARGAYSNARGARTYWSNG-ARGARGARGAYGN 849
Qy 232 GAGATCAGCAGCTGAGAGATGCGCGGTGAGAGTGTGTGCGAGAGCCGAGAGCA 291
Ot D A A P E E N A A C Q E C G A A R D A

DB Y S S E E A E N E E D E D D T E E A E E
Dc 850 TAYWSNWSNARGARCGNARGARAYGARGARGAYACNGARGARGCNGARGAR 909
Qy 292 GAGCCTGAGCAGCAGCTGCGCGAGAGAGCCGAGAGCCGAGAGAG-CAAGAGCA 350
Ot E P E A D A E P V A D E A D E V E X E A

DB D D E E
Dc 910 GAGGARGARGARG 922
Qy 351 GCTGAGAGAAAGG 363
Ot A A E K

RESULT 14
ID P2CG-HUMAN STANDARD: PRT; 546 AA.
AC 015355;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DE 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE PROTEIN PHOSPHATASE 2C GAMMA ISOFORM (EC 3.1.3.16) (P2CG-GAMMA).
GN PMIC.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-SKELETAL MUSCLE;
RX TRAVIS S.M., WELSH M.J.;
RX MEDLINE: 97420453.
CC FEBS LETT. 412:415-419(1997).
CC -1- CATALYTIC ACTIVITY: A PHOSPHOPROTEIN + H₂O -> A PROTEIN +
CC ORTHOPHOSPHATE (THIS ENZYME IS SERINE/THREONINE SPECIFIC).
CC -1- CORRELATOR: BINDS TWO MAGNESIUM OR MANGANESE IONS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED. MOST ABUNDANT IN TESTIS,
CC SKELETAL MUSCLE, AND HEART.
CC -1- SIMILARITY: BELONGS TO THE P2CG FAMILY.
DR EMBL: Y13936; E323054; -
DR PROSITE: PS01032; P2CG; 1.
KW HYDROLASE; MAGNESIUM; MANGANESE; MULTIGENE FAMILY.
FT METAL 40
FT METAL 40 MANGANESE 1 (BY SIMILARITY).
FT METAL 41 MANGANESE 1 (BY SIMILARITY).
FT METAL 60 MANGANESE 1 AND 2 (BY SIMILARITY).
FT METAL 441 MANGANESE 2 (BY SIMILARITY).
FT METAL 496 MANGANESE 2 (BY SIMILARITY).
FT METAL 496 MANGANESE 2 (BY SIMILARITY).
FT DOMAIN 258 326 ASP/GLU-RICH ACIDIC.
FT DOMAIN 266 272 POLY-GLU.
FT DOMAIN 306 311 POLY-GLU.
FT DOMAIN 539 544 POLY-LYS.
SQ SEQUENCE 546 AA: 59271 MW: E2FD44A2 CRC32;

Query Match 3.5%; Score 161; DB 1; Length 546;
Best Local Similarity 39.8%; Pred. No. 1.07e-03;
Matches 53; Conservative 30; Mismatches 48; Indels 2; Gaps 2;

DB D E A E E E E D S E E C S E E E D G
Dc 791 SNGAYGARGCNGARGARGARGAYSNARGARTYWSNG-ARGARGARGAYGN 849
Qy 232 GAGATCAGCAGCTGAGAGATGCGCGGTGAGAGTGTGTGCGAGAGCCGAGAGCA 291
Ot D A A P E E N A A C Q E C G A A R D A

DB Y S S E E A E N E E D E D D T E E A E E
Dc 850 TAYWSNWSNARGARCGNARGARAYGARGARGAYACNGARGARGCNGARGAR 909
Qy 292 GAGCCTGAGCAGCAGCTGCGCGAGAGAGCCGAGAGCCGAGAGAG-CAAGAGCA 350
Ot E P E A D A E P V A D E A D E V E X E A

DB D D E E
Dc 910 GAGGARGARGARG 922
Qy 351 GCTGAGAGAAAGG 363
Ot A A E K

RESULT 15
ID NUCL-RAT STANDARD: PRT; 712 AA.
AC P13383;
DT 01-JAN-1990 (REL. 13, CREATED)
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE NUCLEOLIN (PROTEIN C23).
GN NCL OR NUC.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 90269607.
RX BOURBON H.-M., AMALRIC F.;
RX GENE 88:187-196(1990).
RN [2]
RP SEQUENCE OF 1-44 FROM N.A.
RX MEDLINE: 89121496.
RX BOURBON H.-M., PRUDHOMME M., AMALRIC F.;
RX GENE 68:73-84(1988).
CC -1- FUNCTION: NUCLEOLIN IS THE MAJOR NUCLEOLAR PROTEIN OF GROWING
CC CHROMATIN AND PRE-RIBOSOMAL PARTICLES. IT INDUCES CHROMATIN
CC DECONDENSATION BY BINDING TO HISTONE H1. IT IS THOUGHT TO PLAY A
CC ROLE IN PRE-RNA TRANSCRIPTION AND RIBOSOME ASSEMBLY.
CC -1- SIMILARITY: CONTAINS 4 RNA RECOGNITION MOTIFS (RNP).
CC -1- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLUS.
DR EMBL: M55022; G205792; -
DR EMBL: M55015; G205793; JOINED.
DR EMBL: M55017; G205792; JOINED.
DR EMBL: M55020; G205792; JOINED.
DR PIR: JH0148; JH0148.
DR PROSITE: PS00030; RNP.1; 3.
KW NUCLEAR PROTEIN; PHOSPHORYLATION; METHYLATION; DNA-BINDING; REPEAT;
KW RNA-BINDING.
FT INT-MET 0
FT DOMAIN 142 167 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 187 215 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 241 274 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 312 317 RNA-BINDING (RNP2) (BY SIMILARITY).
FT DOMAIN 357 357 RNA-BINDING (RNP2) (BY SIMILARITY).
FT DOMAIN 398 403 RNA-BINDING (RNP2) (BY SIMILARITY).
FT DOMAIN 432 439 RNA-BINDING (RNP1) (BY SIMILARITY).
FT DOMAIN 490 495 RNA-BINDING (RNP2) (BY SIMILARITY).
FT DOMAIN 525 532 RNA-BINDING (RNP2) (BY SIMILARITY).
FT DOMAIN 576 581 RNA-BINDING (RNP2) (BY SIMILARITY).
FT DOMAIN 612 619 RNA-BINDING (RNP1) (BY SIMILARITY).
FT DOMAIN 651 702 ARG/GLY/PHE-RICH.
FT DOMAIN 57 134 8 X 8 AA TANDEM REPEATS OF X-T-P-X-K-K-X-X.
FT REPEAT 57 64 1.
FT REPEAT 74 81 2.
FT REPEAT 82 89 3.
FT REPEAT 90 97 4.
FT REPEAT 97 103 5.
FT REPEAT 104 111 6.
FT REPEAT 119 126 7.
FT REPEAT 127 134 8.
SQ SEQUENCE 712 AA: 77016 MW: 442EA9F0 CRC32;

Query Match 3.5%; Score 165; DB 1; Length 712;
Best Local Similarity 46.2%; Pred. No. 3.12e-04;
Matches 30; Conservative 19; Mismatches 16; Indels 0; Gaps 0;

Query Match	3.9%;	Score 180;	DB 11;	Length 248;
Best Local Similarity	39.2%;	Pred. No. 7.46e-06;		
Matches	51;	Conservative	22;	Mismatches 57;
			Indels	0;
			Gaps	0;

[illegible]

```

ID RESULT 3
ID 035852 PRELIMINARY; PRT; 248 AA.
AC 035852;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE PREPRODYNORPHIN (FRAGMENT).
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
[]
RN
RC SEQUENCE FROM N.A.
RC STRAIN-129SVJ; TISSUE-LIVER;
RA SHARIFI N., AMENT M., HOCHGESCHWENDER U.;
RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DBDJ DATA BANKS.
DR EMBL; U64968; G2341058; "-
DR PROSITE; PS01252; OPTOIDS_PRCRCRSOR; 1.
FT NON TER 248 248
SQ SEQUENCE 248 AA; 28009 MW; 793AFAF1 CRC32;

```

Query Match	3.9%;	Score 180;	DB 11;	Length 248;
Best Local Similarity	39.2%;	Pred. No. 7.46e-06;		
Matches	51;	Conservative	22;	Mismatches 57;
			Indels	0;
			Gaps	0;

Dd	E D G G G G C A R K A Y G D Q V G H E D L Y K R Y G G
Dt	G A R G A Y G G N G C A R A Y G G A Y C A T G I N G C A Y G A T Y T T A A R M G T A T A G N
Dc	: : : : : : : : : : : : :
Oy	310 G A G C C T G C G A G T A G A G C C G A T G A T G G A C C A A C A C T C A A A A G C T T T T G C G
Qt	E P V A D E A D E V E Q K Q L Q K K R F G
Dd	G F L R I R P K L K W D N Q A C A R Y G G G
Dt	610 G G T T Y T T R M G M A T H M G C C N A R Y T T A A T G G A A A C A R A A M G T A T A G G N G G N
Oy	: : : : : : : : : : : : :
Qt	370 G G C T T C A T G S G G C C G A G C C G G A A C T T G C C A C C A C C A C C G G T T A G G A G 429
Dd	G F T G A R K S A R K L A N Q K R R S S E

Dt	670	TTYTYTMGM	679
		: : :	
QY	430	TTTATGAGGC	439
Qt		F M R	

RESULT	4	
ID	Q69564	
AC	Q69564.	
PRELIMINARY;		
PRT;	715	AA.

DI 1 NOV-1998 (TREMBLER. 01, CREATED)
 DI 01-NOV-1998 (TREMBLER. 01, LAST SEQUENCE UPDATE)
 DT 01-AUG-1998 (TREMBLER. 07, LAST ANNOTATION UPDATE)
 DE 086, IE2.
 GN 086.
 OS HUMAN HERPESVIRUS-6.
 OS VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE; BETAHERPESVIRINAE
 RN [1]
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-U1102, VARIANT A;
 RC MEDLINE; 95266321.
 RX GONPELS U.A., NICHOLAS J., LAWRENCE G., JONES M., THOMSON B.J.,
 RA MARTIN M.E., EFSTATHIOU S., CRAYTON M., MACAULAY H.A.,
 RA VIROLOGY 209:29-51(1995).
 RN [2]
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-U1102, VARIANT A;
 RC MEDLINE; 95266321.

RX MEDLINE: 90080132.
RA LAWRENCE G.L., CHEE M., CRAXTON M.A., GOMPELS U.A., HONESS R.W.,
RL BARRELL B.G.; 64:287-299(1990).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-U1102, VARIANT A;
RX MEDLINE: 91237802.
RA CHANG C.K., BALACHANDRAN N.;
RL J. VIROL. 65:2884-2894(1991).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-U1102, VARIANT A;
RX MEDLINE: 91333007.
RA TEO I.A., GRIFFIN B.E., JONES M.D.;
RL J. VIROL. 65:4670-4680(1991).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN-U1102, VARIANT A;
RX MEDLINE: 91226542.
RA THOMSON B.J., EFSTATHIOU S., HONESS R.W.;
RL NATURE 351:78-80(1991).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN-U1102, VARIANT A;
RX MEDLINE: 91374590.
RA MARTIN M.E.D., NICHOLAS J., THOMSON B.J., NEWMAN C., HONESS R.W.;
RL J. VIROL. 65:5381-5390(1991).
RN [7]
RP SEQUENCE FROM N.A.
RC STRAIN-U1102, VARIANT A;
RX MEDLINE: 92333249.
RA EFSTATHIOU S., LAWRENCE G.L., BROWN C.M., BARRELL B.G.;
RL J. GEN. VIROL. 73:1661-1671(1992).
RN [8]
RP SEQUENCE FROM N.A.
RC STRAIN-U1102, VARIANT A;
RX MEDLINE: 92148942.
RA GENG Y., CHANDRAN B., JOSEPHS S.F., WOOD C.;
RL J. VIROL. 66:1564-1570(1992).
RN [9]
RP SEQUENCE FROM N.A.
RC STRAIN-U1102, VARIANT A;
RX MEDLINE: 93091236.
RA GOMPELS U.A., CARSS A.L., SUN N., ARRAND J.R.;
RL DNA SEQ. 3:25-39(1992).
RN [10]
RP SEQUENCE FROM N.A.
RC STRAIN-U1102, VARIANT A;
RX MEDLINE: 92260671.
RA NEIPEL F., ELLINGER K., FLECKENSTEIN B.;
RL J. VIROL. 66:3918-3924(1992).
RN [11]
RP SEQUENCE FROM N.A.
RC STRAIN-U1102, VARIANT A;
RX MEDLINE: 92333248.
RA THOMSON B.J., HONESS R.W.;
RL J. GEN. VIROL. 73:1649-1660(1992).
RN [12]
RP SEQUENCE FROM N.A.
RC STRAIN-U1102, VARIANT A;
RX MEDLINE: 93187613.
RA ELLINGER K., NEIPEL F., FOA-TOMASI L., CAMPADELLI-FIUME G.,
RL FLECKENSTEIN B.;
RN [13]
RP SEQUENCE FROM N.A.
RC STRAIN-U1102, VARIANT A;
RX MEDLINE: 93324882.
RA GOMPELS U.A., CARRIGAN D.R., CARSS A.L., ARNO J.;
RL J. GEN. VIROL. 74:613-622(1993).
RN [14]
RP SEQUENCE FROM N.A.
RN [15]
RC STRAIN-U1102, VARIANT A;
RX MEDLINE: 93389439.
RA LIU D.X., GOMPELS U.A., NICHOLAS J., LELLIOTT C.;
RL J. GEN. VIROL. 74:1847-1857(1993).
RN [16]
RP SEQUENCE FROM N.A.
RC STRAIN-U1102, VARIANT A;
RX MEDLINE: 94025558.
RA LIU D.X., GOMPELS U.A., FOA-TOMASI L., CAMPADELLI-FIUME G.;
RL VIROLOGY 197:12-22(1993).
RN [17]
RP SEQUENCE FROM N.A.
RC STRAIN-U1102, VARIANT A;
RX MEDLINE: 93331710.
RA PELLET P.E., SANCHEZ-MARTINEZ D., DOMINGUEZ G., BLACK J.B., ANTON E.,
RA GREENAMOVER C., DAMBAUGH T.R.;
RL VIROLOGY 195:521-531(1993).
RN [18]
RP SEQUENCE FROM N.A.
RC STRAIN-U1102, VARIANT A;
RX MEDLINE: 93323202.
RA PEIFFER B., BERNEMAN Z.N., NEIPEL F., CHANG C.K., TIRWATNAPONG S.,
RA CHANDRAN B.;
RL J. VIROL. 67:4611-4620(1993).
RN [19]
RP SEQUENCE FROM N.A.
RC STRAIN-U1102, VARIANT A;
RX MEDLINE: 95146989.
RA GOMPELS U.A., MACAULAY H.A.;
RL J. GEN. VIROL. 76:451-458(1995).
RN [20]
RP SEQUENCE FROM N.A.
RC STRAIN-U1102, VARIANT A;
RX MEDLINE: 94047392.
RA DEMHURST S., DOLIARD S.C., PELLET P.E., DAMBAUGH T.R.;
RL J. VIROL. 67:7680-7683(1993).
RN [21]
RP SEQUENCE FROM N.A.
RC STRAIN-U1102, VARIANT A;
RX MEDLINE: 94202284.
RA SCHIEME U., NEIPEL F., SCHREINER D., FLECKENSTEIN B.;
RL J. VIROL. 68:2978-2985(1994).
RN [22]
RP SEQUENCE FROM N.A.
RC STRAIN-U1102, VARIANT A;
RX MEDLINE: 94181269.
RA THOMPSON J., CHOUDHURY S., KASHANCHI F., DONIGER J., BERNEMAN Z.,
RA FRENKEL N., ROSENTHAL L.J.;
RL ONCOGENE 9:1167-1175(1994).
RN [23]
RP SEQUENCE FROM N.A.
RC STRAIN-U1102, VARIANT A;
RX MEDLINE: 94167865.
RA ZHOU Y., CHANG C.K., QIAN G., CHANDRAN B., WOOD C.;
RL VIROLOGY 199:311-322(1994).
RN [24]
RP SEQUENCE FROM N.A.
RC STRAIN-U1102, VARIANT A;
RX MEDLINE: 94202288.
RA THOMSON B.J., DEMHURST S., GRAY D.;
RL J. VIROL. 68:3007-3014(1994).
RN [25]
RP SEQUENCE FROM N.A.
RC STRAIN-U1102, VARIANT A;
RX MEDLINE: 94202288.
RA THOMSON B.J., DEMHURST S., GRAY D.;
RL J. VIROL. 68:3007-3014(1994).
RN [26]


```

OY 241 CACCTGAAGGAATGCCGCGTGCAGGAGTGCTGTGCAACCCGAGACGCGAGCCTGAG 300
Dt T * R E C R V S G C G A S P R R A * G
Db S S R A G C S S R A S S R A S S S R D S S R A
Dc 2480 CNGSMNSMGCGWMSWMSNMGNCNMSNMGNCNMSNMGNGAYWMSNMGNG 2539
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 301 GCACATGCACAGCCTGTGCGCAGATGAGGCCGATGAG-GTGAGACACAGACCTGCAGAA 359
R C R C R C R R * G R * X W S C R S S C R K
Qt S S K A S S S R A S S R D S S R A S S R D
Db 2540 CNGSMNSMAAAGCWMSWMSNMGNCNMSNMGNGAYWMSNMGNGWMSNMGNG 2599
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 360 AAGTTTGGGGGCTTCACTGCGGGCCCGAAGTCAAGCCCGAAGTTGGCCCAACGAGCG 419
G L G A S L G P G S Q P G S W P T R S G
Qt S S R M D G A V S M S R A S S S R A S S S R D
Db 2600 AYSMSNMGDAVSWMSNMGNCNMSNMGNCNMSNMGNGAYW 2650
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 420 GTTAGCGAGCTTTATGAGCGCATCTCGGTGCTCGAGCATCAAGCCCA 470
S V S L * G S T W S * A C S Q A

```

```

RESULT 7
ID 006634 PRELIMINARY; PRT: 730 AA.
AC 006634;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-FEB-1997 (TREMBLREL. 02, LAST ANNOTATION UPDATE)
DE TRANS-ACTING TRANSCRIPTIONAL PROTEIN BICP4 (FRAGMENT).
GN BICP4.
NC BOVINE HERPESVIRUS TYPE 1 (STRAIN JUR4 (SUBTYPE 1.1));
OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE; ALPHAHERPESVIRINAE.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94025583.
RA SCHWYTER M., VICER C., MENESE O., FRAEFEL C., PACES V.:
RL VIROLOGY 197;349-357(1993).
DR EMBL; L14321; G291538;
KW TRANSCRIPTION REGULATION; TRANS-ACTING FACTOR; ACTIVATOR; ZINC-FINGER;
KW DNA-BINDING; EARLY PROTEIN; REPRESSOR.
FT NON TER 730
SO SEQUENCE 730 AA; 72965 MW; 61B3BE42 CRC32;

```

```

Query Match 3.8%; Score 175; DB 14; Length 730;
Best Local Similarity 43.4%; Pred. No. 3.62e-05;
Matches 36; Conservative 21; Mismatches 26; Indels 0; Gaps 0;

```

```

Db 2108 ARGNGARGAGYGANGARGCNGARGARAYGNGARGAYGARGARGCGNGARGARGARG 2167
Dt A E D E E A E E D G E D E E A E E E E
OY 278 AAGCCCGAGCAGCAGCTGAGCGACATGAGGCGCTGCGCAGATGAGGCGGATGAG 337
A R D A E P E A D A E P V A D E A D E V
Qt A R D A E P E A D A E P V A D E A D E V
Db E E K E E E E
Dt 2168 ARGARGAARAAARGARGARGAR 2190
OY 338 TGGAGCAGAACAGCTGCAGAAA 360
E Q K Q L Q K

```

```

RESULT 8
ID 014550 PRELIMINARY; PRT: 1482 AA.
AC 014550;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE ZINC-FINGER DNA-BINDING PROTEIN.
OS HOMO-SAPIENS (HUMAN).
OC EUKARYOTA; METAEOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
CC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=KIDNEY;

```


Query Match	3.6%	Score 167	DB 11	Length 1117
FT	CONFLICT	664	S -> T (IN REF. 3).	
FT	CONFLICT	744	E -> L (IN REF. 3).	
FT	CONFLICT	822	T -> A (IN REF. 3).	
FT	CONFLICT	1062	MISSING (IN REF. 3).	
SEQUENCE	1117	AA, 12264	MM, 78PB4707	CRC32;

Release 3.1A John F. Collins, Bloccomputing Research Unit
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

```
Run on:      Fri Apr 16 14:34:23 1999;      MasPar time 143.72 Seconds
Tabular output not generated.              1457.286 Million cell updates/sec
```

```

Title: >US-09-011-797-1
Description: (1-932) from US09011797.seq
Perfect Score: 4660
N.A. Sequence: 1 CTCCTCAGCGGTGTCAGCAG.....AAAAAAAAAAAAAGAAATTC 933
Comp: GAAAGTCCACACAAGTCGTC.....TTTTCCTTTTCCTTAAG

```

Scoring table: TABLE bktranslated2

```
Nmatch      STD :      Dbase 0;      Query 0
```

Searched: 116695 seqs, 112361730 bases x 2

Listing first 45 summaries

Database: plr58

Statistics: Mean 73.474; Variance 156.602; scale 0.469

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

Result	Score	Query Match	Length	ID	Description	Pred. No

20	161	3.5	199	2	DE54831	hypothetical protein	1.00e-02
21	165	3.5	254	1	DBH	beta-neoendorphin / d	3.43e-03
22	165	3.5	712	2	UH0148	nucleolin - rat	3.43e-03
23	160	3.4	215	1	S01947	nonhistone chromosoma	1.31e-02
24	160	3.4	215	2	A28897	nonhistone chromosoma	1.31e-02
25	157	3.4	237	1	HADEIV	hemoglobin alpha chain	2.88e-02
26	160	3.4	249	2	S43309	probable HLA class II	1.31e-02
27	158	3.4	254	1	PUBYS	cold shock protein TI	2.22e-02
28	160	3.4	300	2	S54717	probable transcriptio	1.31e-02
29	160	3.4	387	2	S02708	tropoin T - fruit fl	1.31e-02
30	160	3.4	396	2	S13251	tropoin T - fruit fl	1.31e-02
31	160	3.4	1110	2	IS1116	NR-180 - sea lamprey	1.31e-02
32	152	3.3	180	2	JM0387	somatotropin - sei wh	1.04e-01
33	156	3.3	203	2	S60440	probable membrane pro	3.73e-02
34	155	3.3	216	2	S29857	nonhistone chromosoma	4.84e-02
35	156	3.3	457	2	S39713	aldehyde dehydrogenas	3.73e-02
36	155	3.3	587	2	UC5300	Ran GTPase activator	4.84e-02
37	155	3.3	678	2	AS4514	glutamic acid-rich pr	4.84e-02
38	155	3.3	707	2	AS5804	nucleolin - human	4.84e-02
39	155	3.3	758	2	S65169	hypothetical protein	1.04e-01
40	152	3.3	1077	2	S66842	hypothetical protein	1.04e-01
41	156	3.3	3178	2	S42012	6-deoxyerythronolide	3.73e-02
42	156	3.3	3178	2	S13595	6-deoxyerythronolide	3.73e-02
43	150	3.2	247	2	S19334	leucine-rich acidic n	1.35e-01
44	151	3.2	266	2	GA1715	ribosomal protein B2	1.35e-01
45	150	3.2	358	2	S60892	nucleosome assembly p	1.73e-01

hypothetical protein beta-neoendorphin / d	1.00e-02
nucleolin - rat	3.43e-03
nonhistone chromosoma	3.43e-03
nonhistone chromosoma	1.31e-02
hemoglobin alpha chain	1.31e-02
probable HLA class II	2.88e-02
cold shock protein TI	1.31e-02
probable transcriptio	2.22e-02
troponin T - fruit fl	1.31e-02
troponin T - fruit fl	1.31e-02
NF-180 - sea lamprey	1.31e-02
somatotropin - sei wh	1.04e-01
probable membrane pro	3.73e-02
nonhistone chromosoma	4.84e-02
aldehyde dehydrogens	3.73e-02
gran GTPase activator	4.84e-02
glutamic acid-rich pr	4.84e-02
nucleolin - human	4.84e-02
hypothetical protein	1.04e-01
hypothetical protein	1.04e-01
6-deoxyerythronolide	3.73e-02
leucine-rich acidic n	1.31e-01
ribosomal protein HS2	1.35e-01
nucleosome assembly p	1.73e-01

RESULT 1

DATE _____

REFERENCE

Jour

并

非 独

Best I

2

Dt

QY

2

De .

Qy
Q+e

Ph

DE 12

12
of

DB	ENTRY	TITLE	ALTERNATE_NAMES	ORGANISM	DATE	ACCSSIONS	REFERENCE	#authors	#Journal	#Title	#accession	#molecule_type	#residues	#cross-references	COMMENT	KEYWORDS	SUMMARY	Query Match	Best Local Similarity	Matches	277; Conservative	133; Mismatches	93; Indels	0; Gaps	0;
Dt	181	CONCNCNAYCCNARYITNACNWSGCGCTTTTATYACNSNARCNMSNARATGCA																							
Oy	181	CTGCTTACGAGGCTGACGCTGCTGCTTACACGTCGAAAGCTCGAGATGCA																							
Qt		P A D P E L T S A A L Y Q S K A S E M Q																							
Dd	241	CAYTNABNGNATGCCNMGNNGMWSMGTTGTCARGCNMGNCATYCNARCCNGAR																							
Dt	241	CACCTGAGGAATGCGCGCTGTGAGAGAGTGTGTCTCAAGCCCGAGACCGACGCTGAG																							
Oy	241	H L K R M P R V R S V V Q A R D A E P E																							
Qt		A D A E P V A D E A D E V E Q K O L Q K																							
Dd	301	GCAGATGCAAGCTGTGTCAGATGAGCGGAGAGGTGAGACCAAGAGGTGAGAAA																							
Oy	301	A D A E P V A D E A D E V E Q K O L Q K																							
Qt		R F G G F T G A R K S A R K L A N Q K R																							
Dd	361	MGATTTGNGNTTTCACNGNCGMGNABRMSGCMGMAATYTGCNATCARARNGN																							
Dt	361	AGTTTGGGGGCTTCACTGAGGCCCGGAGAGTCAAGCCCGGAGATGTCGAGAAA																							
Oy	361	R F G G F T G A R K S A R K L A N Q K R																							
Qt		F S E F M R Q Y L V L S M Q S S Q R R R																							
Dd	421	TTTWSGARTTTCGAGNCARTATYTYNGTNTSMATGTCARNSWMCARNGMNGN																							
Oy	421	TTGAGTGAAGTTTATGAGGAGTACCTGCTCGACCATGTCACTCAAGCCAAAGCCGCGC																							
Qt		F S E F M R Q Y L V L S M Q S S Q R R R																							
Dd	481	ACNTYTCAYCARAAYGNAAYGT 503																							
Oy	481	ACTCTGACCAGCATGTATGT 503																							
Qt		T L H Q N G N																							
RESULT	2	JC6151	#type complete	orphanin FQ precursor - rat																					
ENTRY		JC6151	#type complete	orphanin FQ precursor - rat																					
TITLE		JC6151	#type complete	orphanin FQ precursor - rat																					
ALTERNATE_NAMES		JC6151	#type complete	orphanin FQ precursor - rat																					
ORGANISM		JC6151	#type complete	orphanin FQ precursor - rat																					
DATE		JC6151	#type complete	orphanin FQ precursor - rat																					
ACCSSIONS		JC6151	#type complete	orphanin FQ precursor - rat																					
REFERENCE		JC6151	#type complete	orphanin FQ precursor - rat																					
#authors		JC6151	#type complete	orphanin FQ precursor - rat																					

Db	PC	A	C	M	C	G	S	F	N	L	K	L	C	I	L	O	C	E	K	V	F
Db	100	C	A	M	C	G	S	F	N	L	K	L	C	I	L	O	C	E	K	V	F
Qt																					
Qt	61	C	G	G	C	T	C	G	G	A	G	C	T	C	T	A	C	T	G	A	A
Qt		P	A	P	G	S	F	N	L	K	L	C	I	L	O	C	E	K	V	F	
Db		P	R	P	L	M	T	L	C	T	K	A	M	A	S	D	S	E	O	L	S
Db	160	C	M	M	G	C	A	C	C	A	T	T	G	A	C	T	T	G	A	C	A
Qt																					
Qt	121	C	C	C	C	C	C	C	T	C	T	G	A	C	T	T	G	A	C	A	A
Qt		P	R	P	L	W	T	L	C	T	K	A	M	A	S	D	S	E	O	L	S
Db		P	A	D	P	E	L	T	S	A	A	L	Y	Q	S	K	A	S	E	M	O
Db	220	C	M	G	C	A	C	C	A	T	T	G	A	C	T	T	G	A	C	A	A
Qt																					
Qt	181	C	C	G	C	G	A	T	C	C	A	C	T	C	A	C	T	C	T	T	T
Qt		P	A	D	P	E	L	T	S	A	A	L	Y	Q	S	K	A	S	E	M	O
Db		H	L	K	R	M	P	R	V	R	S	V	V	Q	A	R	D	A	E	P	E
Db	280	C	A	T	T	G	A	C	C	A	T	T	G	A	C	T	T	G	A	C	A
Qt																					
Qt	241	C	A	C	C	G	A	G	A	G	A	T	C	C	C	C	G	T	G	A	G
Qt		H	L	K	R	M	P	R	V	R	S	V	V	Q	A	R	D	A	E	P	E
Db		A	D	A	E	P	V	A	D	E	V	E	V	Q	A	R	D	A	E	P	E
Db	340	G	G	N	G	A	T	G	A	G	A	T	C	C	C	C	G	T	G	A	G
Qt																					
Qt	301	G	A	G	A	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	
Qt		A	D	A	E	P	V	A	D	E	V	E	V	Q	A	R	D	A	E	P	E
Db		R	F	G	G	F	T	G	A	R	K	S	A	R	K	L	A	N	Q	K	R
Db	400	M	E	N	T	T	G	A	C	C	A	T	T	G	A	C	C	A	T	T	G
Qt																					
Qt	361	A	G	G	T	T	G	G	G												


```

#title      Structure and regional distribution of nociceptin/orphanin Fy
#accession  J04652
#molecule_type mRNA
#residues 1-187 ##label HOU
#cross-references DBS:D82866; NID:g1311472; PID:d1012281; PID:g1311473
#experimental_source brain
#note      The authors translated the codon CTC for residue 72 as
            Lys and GAC for the opiate precursor family, and functions
            as a neuropeptide precursor and plays a role in neuronal
            differentiation and development.

GENETICS
#gene
KEYWORDS
FEATURE
1-18
19-187
#domain signal sequence #status predicted #label SIG\
#product neuropeptide precursor, N23K #status predicted
            neuropeptide
N23K
brain; differentiation; G protein-coupled receptor;
109-126
141-157
160-176
SUMMARY
#length 187 #molecular-weight 20864 #checksum 8758

Query Match
Best Local Similarity 51.8%; Pred. No. 4,776-200;
Matches 270; Conservative 123; Mismatches 110; Indels 18; Gaps 8;

Db      L S S V F S S C P R D C L T C O E K L H
40 YTNNSNGTNTTNTTWSNMTGTCGGMGAGTGYTNACNTGTCARARARTNTY 99
   : : : : : : : : : : : : : : : : : : : : : : : :
1 CTCTCAGCGGTGTACACAGCTGCCGAGAGTGCCTACCTGCGAGAGGCTTCAC 60
   L S S V F S S C P E D C L T C O E K V F
Db      P A P D S F N L K T C I L O C E E K V F
100 CCGNCNGNMGATNTTAAATYTNARACTGTATHTYTCARTGTGACARARARTNTY 159
   : : : : : : : : : : : : : : : : : : : : : : : :
61 CCGGCTCGGGACCTTCAACCTGAACCTGTGATCTCTCAGTGTGAAGAGGTTCTC 120
   P A P G S F N L K L C I L O C E E K V F
Db      P R P L W T V C T K V M A S G S G O L S
160 CCGNCGNTNTTNGACNGTNGTCAAGATGATGACNNSNGNCGNCAARTNNSN 219
   : : : : : : : : : : : : : : : : : : : : : : : :
121 CCGCGCTCTGTGACTCTTGGACCAAGCATGCGAGCTGACGAGCTCAGC 180
   P R P L W T L C T K A M A S D S E M O
Db      P A D P E L V S A L Y O P K A S E E M O
220 CCGNCNGVCCNGARYNTNGTWSNGCNVYTNAYCARCNAARCSNARATGAC 279
   : : : : : : : : : : : : : : : : : : : : : : : :
181 CCTCTGTATCAGAGCTCAGCTCGCGTCTTACAGCGAAGACCTCGGAGATGAC 240
   P A D P E L T S A A L Y O S K A S E M O
Db      H L K R M P R V R S L V Q V R D A E P C
280 CAYTINARMGNAATGCCNMGTNGMWSYTNLTNCAKTNTNGNAYCNCARCSNGN 339
   : : : : : : : : : : : : : : : : : : : : : : : :
241 CACCTGAAGAAAGCGCGGTGTGAG--GA---GT---GT---G-G-TGCA-AGCCG-- 284
   H L K R M P R V R V X V V V X X A X P X
Db      A D A E P G A D A E P G A D D A E E V E
340 GCGNAGYCGNARCCNGGNGCNGAGCNGARCSNGGNCNGAYAYCNCARARATNGAR 399
   : : : : : : : : : : : : : : : : : : : : : : : :
285 --AGACGAGAGCCTGAGCGAGATGACAGACCTGTGCGAGATAGGCCCATGAGGGAG 342
   x D A E P E A D A E P V A D E A D E V E
Db      Q K O L O K R F G G F T G A R K S A R K
400 CARARCARRYNCCARARMGNTTYGNGGNTTYACNGNCCNNGNARNSNCGNNAAR 459
   : : : : : : : : : : : : : : : : : : : : : : : :
343 CAGAGCGAGCGCAAGAAAGTGTGGGGGCTTCTGAGCGGCGGAAGCAGCGGAG 402
   Q K O L O K R F G G F T G A R K S A R K
Db      L A N O K R F S E F M R O Y L V L S M O
460 YTNCGNAAVCARARMGNTTWSGARTTYATGNCNARAYTNTNTNNSNATGAC 519

```

[illegible]


```

#note      cross-linking
            the six prolines at the third position of the tripeptide
            repeating unit (G-X-Y) may be hydroxylated
CLASSIFICATION #superfamily collagen alpha 1(I) chain; fibrillar collagen
                carboxyl-terminal homology; von Willebrand factor type C
                repeat homology
KEYWORDS      coiled coil; extracellular matrix; glycoprotein; trimer;
              triple helix
FEATURE       #modified_site allysine (lys) #status predicted
              #length 53 #checksum 1461
SUMMARY
Insufficient memory to reconstruct alignment.

RESULT      8
ENTRY       CGRTIS #type fragments
TITLE       collagen alpha 1(I) chain - rat (fragments)
ORGANISM    #formal_name Rattus norvegicus #common_name Norway rat
DATE        13-Jul-1981 #sequence_revision 13-Jul-1981 #text_change
            19-Oct-1995
ACCESSIONS  A90559; A90552; A90353; A90566; A90357; A90362;
            A90379; A91209; A91385; A02854; A02855
REFERENCE   #authors      Bornstein, P.
            #journal      Biochemistry (1969) 8:63-71
            #title         Comparative studies of rat skin and tendon collagen.
            #note          I. The absence of a short sequence at the amino terminus
            of the skin alpha1 chain.
            #cross-references MUID:69155173
            #contents      CNB80 and CNB81
            #accession     A90559
            #molecule_type protein
            #residues      1-19 #label B01
            #experimental_source tendon
            #note          sequences from skin and tendon appear to be identical
            the amino-terminal tetrapeptide may be removed by
            limited proteolysis during extraction
REFERENCE   #authors      Kang, A.H.; Bornstein, P.; Piez, K.A.
            #journal      Biochemistry (1967) 6:788-795
            #title         The amino acid sequence of peptides from the cross-linking
            region of rat skin collagen.
            #cross-references MUID:67162268
            #contents      CNB81
            #accession     A90552
            #molecule_type protein
            #residues      5-19 #label KAN
            #experimental_source skin
REFERENCE   #authors      Bornstein, P.
            #journal      J. Biol. Chem. (1967) 242:2572-2574
            #title         The incomplete hydroxylation of individual prolyl residues in
            collagen.
            #cross-references MUID:67165368
            #contents      CNB82
            #accession     A92029
            #molecule_type protein
            #residues      20-55 #label B02
            #experimental_source skin and tendon
REFERENCE   #authors      Butler, W.T.; Ponds, S.L.
            #journal      Biochemistry (1971) 10:2076-2081
            #title         Chemical studies on the cyanogen bromide peptides of rat skin
            collagen. Amino acid sequence of alpha1-CB4.
            #cross-references MUID:71263178
            #contents      CNB84
            #accession     A90353
            #molecule_type protein
            #residues      56-102 #label B01
            #experimental_source skin
REFERENCE   #authors      Butler, W.T.
            #note          A90566

```

```

#journal    Biochemistry (1970) 9:44-50
#title      Chemical studies on the cyanogen bromide peptides of rat skin
            collagen. The covalent structure of alpha1-CB5, the major
            hexose-containing cyanogen bromide peptide of alpha1.
            #cross-references MUID:70085124
            #contents      CNB85
            #accession     A90566
            #molecule_type protein
            #residues      103-139 #label B02
            #experimental_source skin
REFERENCE   #authors      Balian, G.; Click, E.M.; Bornstein, P.
            #journal      Biochemistry (1971) 10:4470-4478
            #title         Structure of rat skin collagen alpha1-CB8. Amino acid
            sequence of the hydroxylamine-produced fragment HA1.
            #cross-references MUID:72136131
            #contents      CNB88
            #accession     A90357
            #molecule_type protein
            #residues      140-238 #label B01
            #experimental_source skin
REFERENCE   #authors      Balian, G.; Click, E.M.; Hermanson, M.A.; Bornstein, P.
            #journal      Biochemistry (1972) 11:3798-3806
            #title         Structure of rat skin collagen alpha1-CB8. Amino acid
            sequence of the hydroxylamine-produced fragment HA2.
            #cross-references MUID:73006942
            #contents      CNB88
            #accession     A90362
            #molecule_type protein
            #residues      239-418 #label B02
            #experimental_source skin
REFERENCE   #authors      Butler, W.T.; Underwood, S.P.; Finch Jr., J.E.
            #journal      Biochemistry (1974) 13:2946-2953
            #title         Chemical studies on the cyanogen bromide peptides of rat skin
            collagen. Amino acid sequence of alpha1-CB3.
            #cross-references MUID:74271984
            #contents      CNB83
            #accession     A90379
            #molecule_type protein
            #residues      419-567 #label B03
            #experimental_source skin
REFERENCE   #authors      Stoltz, M.; Timpl, R.; Furtmayr, H.; Kuehn, K.
            #journal      Eur. J. Biochem. (1973) 37:287-294
            #title         Structural and immunogenic properties of a major antigenic
            determinant in neutral salt-extracted rat-skin collagen.
            #cross-references MUID:74011954
            #contents      CNB86
            #accession     A91209
            #molecule_type protein
            #residues      568-651 #label S01
            #experimental_source skin
            #note          this region probably corresponds to positions 949-1032
            of the alpha 1(I) chain
            the major antigenic determinant (of neutral
            salt-extracted rat skin collagen) involves at least
            Tyr-Asp at positions 650-651 above and probably the
            nine residues that precede the dipeptide
            #note          A91385
            #authors      Stoltz, M.; Timpl, R.; Kuehn, K.
            #journal      FEBS Lett. (1972) 26:61-65
            #title         Non-helical regions in rat collagen alpha1-chain.
            #cross-references MUID:73049495
            #contents      CNB86
            #accession     A91385
            #molecule_type protein
            #residues      651-671 #label S02
            #experimental_source skin
            #note          the composition of peptides comprising residues 1-9 and
            1-19 confirms the sequence
            this region (residues 651-671 above) probably

```



```

**experimental source skin
REFERENCE A90379
#authors Butler, W.T.; Underwood, S.P.; Finch Jr., J.E.
#journal Biochemistry (1974) 13:2946-2953
#title Chemical studies on the cyanogen bromide peptides of rat skin
#cross-references MUID:74271984
#contents CNBr3
#accession A90379
#molecule_type protein
#residues 419-567 ##label BU3
#experimental_source skin
REFERENCE A91209
#authors Stoltz, M.; Timpl, R.; Furchmayr, H.; Kuehn, K.
#journal Eur. J. Biochem. (1973) 37:287-294
#title Structural and immunogenic properties of a major antigenic
#cross-references MUID:74011954
#contents CNBr6
#accession A91209
#molecule_type protein
#residues 568-651 ##label ST1
#experimental_source skin
#note this region probably corresponds to positions 949-1032
#note of the alpha 1(I) chain
#note the major antigenic determinant (of neutral
salt-extracted rat skin collagen) involves at least
Tyr-Asp at positions 650-651 above and probably the
nine residues that precede the dipeptide

REFERENCE A91385
#authors Stoltz, M.; Timpl, R.; Kuehn, K.
#journal FEBS Lett. (1972) 26:61-65
#title Non-helical regions in rat collagen alpha1-chain.
#cross-references MUID:73049495
#contents CNBr6
#accession A91385
#molecule_type protein
#residues 651-671 ##label ST2
#experimental_source skin
#note the composition of peptides comprising residues 1-9 and
1-19 confirms the sequence
this region (residues 651-671 above) probably
corresponds to positions 1032-1052 of the alpha 1(I)
chain
COMMENT Prolines and lysines at the third position of the tripeptide
repeating unit (G-X-Y) are hydroxylated to varying extents.
Prolines are predominately 4-hydroxylated. Lysines are
5-hydroxylated and subsequently O-glycosylated.
COMMENT The order of the nine CNBr peptides in the alpha 1(I) chain of rat
skin collagen was determined as 0,1,2,4,5,8,3,7,6.
COMMENT The complete chain contains 1052 residues.
CLASSIFICATION #superfamily collagen alpha 1(I) chain; fibrillar collagen
carboxyl-terminal homology; von Willebrand factor type C
repeat homology
KEYWORDS coiled coil; extracellular matrix; glycoprotein; trimer;
triple helix
FEATURE
1 #modified_site blocked amino end (Glx) (probably
pyrrolidone carboxylic acid) #status experimental\
9 #modified_site allysine (Lys) #status experimental\
103,424,547 #binding_site carbohydrate (Lys) (covalent) #status
experimental\
103 #modified_site 5-hydroxylysine (Lys) #status
experimental\
424,547 #modified_site 5-hydroxylysine (Lys) (partial) #status
experimental
SUMMARY #length 671 #checksum 936
Insufficient memory To Reconstruct alignment.
RESULT 9
ENTRY PC4356 #type fragment
ENTRY PC4356 #type fragment

```

```

TITLE prodyndorphin - mouse (fragment)
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 02-Sep-1997 #sequence_revision 05-Sep-1997 #text_change
26-Feb-1998
ACCESSION PC4356
#authors Gulya, K.; Orpana, A.K.; Sikela, J.M.; Hoffman, P.L.
#journal Mol. Brain Res. (1993) 20:1-8
#title Prodyndorphin and vasopressin mRNA levels are differentially
affected by chronic ethanol ingestion in the mouse.
#accession PC4356
#molecule_type DNA
#residues 1-63 ##label GUL
#accession PC4357
#molecule_type protein
#residues 1-7:57-63 ##label GU2
#experimental_source brain
CLASSIFICATION #superfamily proenkephalin
FEATURE
1-4 #product alpha-neoendorphin #status predicted #label
ANE\
32-36,51-55 #product leucine endorphalin #status predicted #label
LEE\
37-48 #product dynorphin-A #status predicted #label DNA\
56-63 #product dynorphin-B #status predicted #label DNB
SUMMARY #length 63 #checksum 8609
Query Match 3.9%; Score 180; DB 2; Length 771;
Best Local Similarity 39.2%; Pred. No. 5,24e-05;
Matches 51; Conservative 22; Mismatches 57; Indels 0; Gaps 0;

Db E D G G Q D G D Q V G H E D L Y K R Y G
Dt 40 GARAYGNGNGNCARAYGNGAYCARGTNGCNCAYARGAYVNTYVAAAMGTAYGNGN 99
OY 11: 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Qt 310 GAGCCTGCGCAGATGAGCGCGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 369
Dt E P V A D E A D E V E Q K Q L Q K R F G
Db G F L R R I R P K L K W D N Q K R Y G G
Dt 100 GGNTRYTNMGNMGNMNCNCAARTNARTGAGAYVCAARAARMTAYGNGNGN 159
OY 11: 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Qt 370 GCGTCTAGTGGCGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 429
Dt G F T G A R K S A R K L A N Q K R F S E
OY 430 TTTATGAGGC 439
Qt F M R

RESULT 9
ENTRY PC4356 #type fragment
ENTRY prodyndorphin - mouse (fragment)
TITLE #formal_name Mus musculus #common_name house mouse
ORGANISM 02-Sep-1997 #sequence_revision 05-Sep-1997 #text_change
26-Feb-1998
ACCESSION PC4356
#authors Gulya, K.; Orpana, A.K.; Sikela, J.M.; Hoffman, P.L.
#journal Mol. Brain Res. (1993) 20:1-8
#title Prodyndorphin and vasopressin mRNA levels are differentially
affected by chronic ethanol ingestion in the mouse.
#accession PC4356
#molecule_type DNA
#residues 1-63 ##label GUL
#accession PC4357
#molecule_type protein
#residues 1-7:57-63 ##label GU2
#experimental_source brain
CLASSIFICATION #superfamily proenkephalin
FEATURE
1-4 #product alpha-neoendorphin #status predicted #label
ANE\

```



```

OY      215 ACCAGTCGAAAGCCTCGCATGACACGACCTGAGAGAAAGCCCGCTGATGAGAGTGTGG 274
      P V E S L G D A A P E E N A A C Q E V V
Db      812 ARGARGARGARGARGARGARGARGARGARGARGARGARGARGARGARGARGARGARG 871
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      275 TGCAGCCCGACGACCGACGACCTGAGGCAATGACAGCGCTGCGCATGAGAGCCGATG 334
      Q A R D A E P E A D A E P V A D E A D E
Db      872 AYGARGNGARGARGARGARGARGARGARGARGARGARGARGARGARGARGARGARG 888
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      335 AGCTGAGCAGAGCAG 351
      V E Q K Q

RESULT  12
ENTRY   VIGH2      #type complete
TITLE   vitellogenin II precursor - chicken
CONTAINS #formal_name Gallus gallus #common_name chicken
ORGANISM #formal_name Gallus gallus #sequence_revision 30-Jun-1988 #text_change
DATE     28-Aug-1985
ACCESSIONS
#authors A92941; A93502; C29184; I50440; S07788; A03335; A23177;
#molecule_type DNA
#journal van het Schip, F.D.; Samalio, J.; Broos, J.; Ophuis, J.;
#title   Mojot, M.; Gruber, M.; AB, G.
          J. Mol. Biol. (1987) 196:245-260
          Nucleotide sequence of a chicken vitellogenin gene and
          derived amino acid sequence of the encoded yolk precursor
          protein.
#cross-references M01D:88011328
#accession A92941
#residues 1-1650 ##label VAN
#cross-references GB:X13607; NID:g63886; PID:g63887
#note     1840-Ale was also found

REFERENCE
#authors A93502
#journal Burch, J.B.E.
#title   Nucleic Acids Res. (1984) 12:1117-1135
          Identification and sequence analysis of the 5' end of the
          major chicken vitellogenin gene.
#cross-references M01D:84118805
#accession A93502
#molecule_type DNA
#residues 1-71 ##label BUR
#cross-references GB:X00345; NID:g63872; PID:g63873

REFERENCE
#authors A91754
#journal Clark, R.C.
#title   Int. J. Biochem. (1985) 17:983-988
          The primary structure of avian phosphatins. Contributions
          through the Edman degradation of methylmercaptovinins
          prepared from the constituent phosphoproteins.
#cross-references M01D:86056531
#accession C29184
#molecule_type protein
#residues 1112-1188 ##label CLA

REFERENCE
#authors I50440
#journal Byrne, B.M.; van Het Schip, A.D.; van de Klundert, J.A.
#title   Biochemistry (1984) 23:4275-4279
          Amino acid sequence of phosphatins derived from the nucleotide
          sequence of part of the chicken vitellogenin gene.
#cross-references M01D:85023316
#accession I50440
#status   translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 1092-1338 ##label BYR
#cross-references GB:K02113; NID:g212878; PID:g212879
REFERENCE
#authors S07788
#journal Phillipsen, J.N.J.; de Vries, J.E.; Samalio, J.; van Dijk, C.;
#title   Arndberg, A.C.; AB, G.

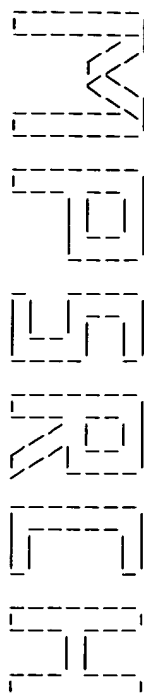
```

```

#journal J. Mol. Evol. (1989) 28:185-190
#title   Characterization of a polymorphism in the 3' part of the
          chicken vitellogenin gene.
#cross-references M01D:89178745
#accession S07788
#status   translation not shown
#molecule_type DNA
#residues 1705-1757 ##label ABG
#cross-references EMBL:X14729; NID:g63881; PID:e13369; PID:g1334751
#note     The source is designated as Gallus domesticus
          Vitellogenin is synthesized in the liver of oviparous vertebrates
          in response to steroid (estrogen) induction.
          Phosvitin, an egg yolk storage protein, is one of the most highly
          phosphorylated (10%) proteins in nature.

GENETICS
#gene
#introns 14/1; 21/1; 71/3; 155/3; 209/3; 259/2; 309/3; 362/2; 417/3;
          459/3; 532/3; 605/2; 638/1; 679/3; 713/3; 753/3; 817/1;
          848/2; 910/2; 958/3; 1033/1; 1091/3; 1321/3; 1338/5;
          1368/3; 1439/3; 1487/2; 1533/3; 1579/1; 1641/2; 1672/1;
          1704/3; 1758/1; 1812/1
CLASSIFICATION #superfamily vitellogenin
KEYWORDS egg yolk; glycoprotein; liver; phosphoprotein
FEATURE
1-15
112-1328
1123,1124,1125,
1126,1127,1128,
1130,1131,1135,
1136,1137,1138,
1139,1140,1142,
1143,1167,1168,
1169,1170,1172,
1173,1175,1176,
1178,1179,1182,
1183,1184,1186,
1187
1189,1190,1193,
1194,1195,1196,
1197,1198,1199,
1200,1201,1202,
1203,1204,1206,
1207,1208,1209,
1210,1211,1212,
1213,1214,1215,
1217,1219,1220,
1221,1222,1223,
1224,1226,1227,
1228,1229,1230,
1231,1233,1235,
1236,1237,1239,
1240,1241,1242,
1243,1244,1245,
1246,1247,1248,
1249,1250,1251,
1252,1254,1255,
1256,1257,1259,
1260,1261,1262,
1263,1264,1266,
1267,1268,1271,
1273,1276,1282,
1283,1284,1285,
1286,1287,1288,
1289,1291,1293,
1296
1280
1301,1308,1309,
1310,1311,1312,
1313,1314,1315
#binding_site phosphate (Ser) (covalent) #status
experimental\
#binding_site phosphate (Ser) (covalent) #status
experimental\
#binding_site phosphate (Ser) (covalent) #status
predicted
#binding_site phosphate (Ser) (covalent) #status
predicted

```

(TM)

Release 3.1a John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Run on: Tue Apr 27 11:04:21 1999; Maspar time 1255.89 Seconds
1328.940 Million cell updates/sec

Tabular output not generated.

Title: >US-09-011-797-1
Description: (1-932) From US09011797.seq
Perfect Score: 932
N.A. Sequence: 1 CTTCTCCAGCGTTCAGCAG.....AAAAAAGGAATTC 932
Comp: GAGAGGTCGCACAGTCGTC.....TTTTTTTTTTCTTAAAG

Scoring table: TABLE default
Gap 6

Match STD : Dbase 0; Query 0

Searched: 2275026 seqs, 895388244 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database:

embl-est36
1:em-est1 2:em-gss1 3:em-gss2 4:em-gss3
genbank-est109

5:gb-est1 6:gb-est10 7:gb-est11 8:gb-est12 9:gb-est13
10:gb-est14 11:gb-est15 12:gb-est16 13:gb-est17
14:gb-est18 15:gb-est19 16:gb-est2 17:gb-est20
18:gb-est21 19:gb-est3 20:gb-est4 21:gb-est5 22:gb-est6
23:gb-est7 24:gb-est8 25:gb-est9 26:gb-gss1 27:gb-gss2
28:gb-gss3 29:gb-gss4

Statistics: Mean 11.547; Variance 3.810; Scale 3.031

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match	Length	DB	ID	Description	Pred. No.
C 1	377	40.5	382	15	AI044049	UI-R-C1-jv-f-08-0-UI.s	0.00e+00	
C 2	355	38.1	621	18	AI178977	EST1222659 Normalized r	0.00e+00	
C 3	311	33.4	374	14	AA801447	EST190944 Normalized r	0.00e+00	
C 4	221	23.7	362	20	W50611	md04d04.r1 Soares mous	2.49e-227	
C 5	193	20.7	372	18	AI153441	uc33c08.r1 Soares 2NDM	1.79e-192	
C 6	176	18.9	249	22	AA023304	mh70d02.r1 Soares mous	1.87e-171	
C 7	165	17.8	445	10	AA648997	ns41g11.s1 NCI_CGAP_GC	3.73e-159	
C 8	162	17.4	319	13	AA845221	ak76d12.s1 Bartshead sp	3.00e-154	
C 9	155	16.6	365	13	AA760677	n209c12.s1 NCI_CGAP_GC	1.10e-145	
C 10	134	16.5	397	8	AA283020	zt16b04.s1 NCI_CGAP_GC	1.83e-144	
C 11	154	16.5	431	8	AA282757	z891a07.s1 NCI_CGAP_GC	1.83e-144	
C 12	154	16.5	431	13	AA806350	oc26f09.s1 NCI_CGAP_GC	1.83e-144	
C 13	151	16.2	344	12	AA809444	ob43e01.s1 NCI_CGAP_GC	8.33e-141	

C 14	146	15.7	305	7	HUM116D04A	Human fetal brain cDNA	1.01e-134
C 15	132	14.2	499	14	AA801448	EST190945 Normalized r	8.90e-118
C 16	131	14.1	364	16	RA51100	y938b10.s1 Homo sapien	1.42e-116
C 17	117	12.6	392	8	AA282862	z591a08.r1 NCI_CGAP_GC	8.26e-100
C 18	114	12.2	186	20	HSAAACBIR	H. sapiens putatively	3.05e-96
C 19	100	10.7	199	20	W77365	m655h09.r1 Soares mous	9.62e-80
C 20	70	7.5	458	8	AA283106	zt16b04.r1 NCI_CGAP_GC	1.69e-45
C 21	59	6.3	352	12	AA754459	y938b10.s1 Homo sapien	1.58e-33
C 22	59	6.3	359	12	RI19874	y938b10.s1 Homo sapien	1.58e-33
C 23	56	6.0	252	12	AA754459	y938b10.s1 Homo sapien	1.58e-33
C 24	50	5.4	247	12	AA754458	y938b10.s1 Homo sapien	1.58e-33
C 25	49	5.3	247	12	AA754458	y938b10.s1 Homo sapien	1.58e-33
C 26	36	3.9	2275	11	AF034173	Homo sapiens ntcon2 co	1.77e-10
C 27	34	3.6	340	13	AA885285	al58g01.s1 Soares NFL	1.12e-08
C 28	31	3.3	2275	11	AF034173	Homo sapiens ntcon2 co	4.49e-06
C 29	29	3.1	267	14	AA873223	oh70g06.s1 NCI_CGAP_K1	2.07e-04
C 30	29	3.1	356	17	AF015060	Mus musculus 2-cell em	2.07e-04
C 31	29	3.1	356	14	AA835988	msc1908.s1 NCI_CGAP_GC	2.07e-04
C 32	29	3.1	452	8	AA414281	vc60a08.s1 Knowles Sol	2.07e-04
C 33	29	3.1	456	18	AA147376	q964c09.s1 Soares test	2.07e-04
C 34	29	3.1	520	6	AA436394	zy44a03.s1 Soares ovar	2.07e-04
C 35	29	3.1	603	8	AA100987	z41e03.s1 Stratagene	2.07e-04
C 36	28	3.0	223	18	AA136193	UI-R-C2pns-d-05-0-UI	1.33e-03
C 37	28	3.0	297	10	AA640712	nr22g05.r1 NCI_CGAP_Pt	1.33e-03
C 38	28	3.0	320	20	W91592	MIA.E03.070.C MTA Adul	1.33e-03
C 39	28	3.0	341	17	AA1088019	ov60e10.x1 Soares NSF	1.33e-03
C 40	28	3.0	406	22	AA204983	z571a03.s1 Stratagene	1.33e-03
C 41	28	3.0	433	14	AA830841	oc54c01.s1 NCI_CGAP_GC	1.33e-03
C 42	28	3.0	504	11	AA682291	z185h06.s1 Soares feta	1.33e-03
C 43	28	3.0	505	15	AA036290	ud60f04.x1 Sugano mous	1.33e-03
C 44	27	2.9	477	13	AA837073	oc18d12.s1 NCI_CGAP_GC	8.20e-03
C 45	27	2.9	622	11	HS069183	Human DRES 44 mRNA seq	8.20e-03

ALIGNMENTS

RESULT 1	AI044049	382 bp	MRNA	EST	06-JUL-1998
LOCUS	UI-R-C1-jv-f-08-0-UI.s1	UI-R-C1 Rattus norvegicus	CDNA clone		
DEFINITION	UI-R-C1-jv-f-08-0-UI 3', mRNA sequence.				
ACCESSION	AI044049				
NID	93290952				
KEYWORDS	EST.				
SOURCE	Norway rat.				
ORGANISM	Rattus norvegicus				
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;				
AUTHORS	Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
TITLE	1 (bases 1 to 382)				
JOURNAL	Bonaldo,M.F., Lennon,G. and Soares,M.B.				
MEDLINE	Normalization and subtraction: two approaches to facilitate gene				
COMMENT	discovery				
	Genome Res. 6 (9), 791-806 (1996)				
	97044477				

Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence tag present in the cDNA between the NotI site and the
oligo-dT track served to identify it as a clone from the normalized
adult brain library. cDNA library Preparation: M. Fatima Bonaldo,
Ph.D. Clone distribution: clones will be available through Research
Genetics
Seq primer: M13 Forward.
Location/Qualifiers
1..382
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/note="vector: pRT33-Pac (pharmacia) with a modified
polylinker: Site_1: Not I; Site_2: Eco RI; the UI-R-C1

library is a subtracted library derived from the UI-R-C0 library, which is a subtracted library derived from the UI-R-A1 and UI-R-E1 libraries. The UI-R-A1 library consisted of a mixture of individually tagged normalized libraries constructed from rat placenta, adult lung, brain, liver, kidney, heart, spleen, ovary, and muscle. The UI-R-E1 library consisted of a mixture of individually tagged normalized libraries constructed from 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dr track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-C1) was constructed as follows: PCR amplified cDNA inserts from UI-R-C0 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the UI-R-C0 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-C1 library. This procedure has been previously described (Bonaldi, Lennon and Soares, Genome Research 6: 791-806, 1996).

/db.xref="taxon:10116"
/clone="UI-R-C1-jv-f-08-0-UI"
/clone.lib="UI-R-C1"
/dev.stage="Adult"
/lab_host="DH10B (Life Technologies)"

BASE COUNT 100 a 105 g 102 t 1 others
ORIGIN

Query Match 40.5%; Score 377; DB 15; Length 382;
Best Local Similarity 99.2%; Pred. No. 0.00e+00;

Matches 379; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 1 TTTTATTTTATTTTACATTTGCTCAGTAGAAGTTAATGAGAAAGTGTTC 60
|||||
Cp 922 TTTTATTTTATTTTACATTTGCTCAGTAGAAGTTAATGAGAAAGTGTTC 863
Db 61 AACATCAGTCAAAAAGACAGCTCCTTTACAACCAATTATGCGAGTGAAGCA 120
|||||
Cp 862 AACATCAGTCAAAAAGACAGCTCCTTTACAACCAATTATGCGAGTGAAGCA 803
Db 121 CCCAGGCTGATTTATTTCTCTCACCCTGCTACGAGAGGCGAGAGATGACAA 180
|||||
Cp 802 CCCAGGCTGATTTATTTCTCTCACCCTGCTACGAGAGGCGAGAGATGACAA 743
Db 181 GAAGATGGGCGCATTAATGTTAGCTGAGGAAATTAACAAATACACTGATGAGCA 240
|||||
Cp 742 GAAGATGGGCGCATTAATGTTAGCTGAGGAAATTAACAAATACACTGATGAGCA 683
Db 241 ACAGATTTGTTGAGACATGCTGTGGGAGAGTCCGCGGTCTTCAAGTCCACAGGAA 300
|||||
Cp 682 ACAGATTTGTTGAGACATGCTGTGGGAGAGTCCGCGGTCTTCAAGTCCACAGGAA 623
Db 301 GGAGGAATCTCTACCTCCGCGCTGACAGTCTTGTGTGACACATGCTTACGCGGGCTC 360
|||||
Cp 622 GGAGGAATCTCTACCTCCGCGCTGACAGTCTTGTGTGACACATGCTTACGCGGGCTC 563
Db 361 ACCTGNATGCTCATGGGTGCA 382
|||||
Cp 562 ACCTGNATGCTCATGGGTGCA 541

RESULT 2
LOCUS A1178977 621 bp mRNA EST 08-OCT-1998
DEFINITION EST222659 Normalized rat spleen, Bento Soares Rattus sp. cDNA clone
ACCESSION RSPBT94.3 end, mRNA sequence.
NID A1178977
KEYWORDS 93729615
SOURCE EST.
Rattus sp.

ORGANISM Rattus sp.
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 621)
AUTHORS Lee,N.H., Gloder,A., Chandra,I., Mason,T.M., Quackenbush,J.,
Kerlanage,A.R. and Adams,M.D.
TITLE Rat Genome Project: Generation of a Rat EST (RST) Catalog & Rat
Gene Index
JOURNAL Unpublished (1998)
COMMENT Other_ESTs: TC52405
Contact: Lee, NH
ATCC
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.

FEATURES
Location/Qualifiers
1..621
/organism="Rattus sp."
/note="Organ: spleen; Vector: pT7T3Pac; Site_1: EcoRI;
Site_2: NotI"
/db.xref="taxon:10118"
/db.xref="RSPBT94"
/clone.lib="Normalized rat spleen, Bento Soares"
BASE COUNT 140 a 166 c 166 g 149 t
ORIGIN

Query Match 38.1%; Score 355; DB 18; Length 621;
Best Local Similarity 98.1%; Pred. No. 0.00e+00;
Matches 362; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 1 ATTAACATTTTGTCTCAGTAGAAGTTAATGAGAAAGTGTTCACAAATCAGTCAAA 60
|||||
Cp 908 ATTAACATTTTGTCTCAGTAGAAGTTAATGAGAAAGTGTTCACAAATCAGTCAAA 849
Db 61 AAGAAGAGCTCCTTTACAACCAATTATGCGAGTGAAGCAAAACCCAGTGTGATT 120
|||||
Cp 848 AAGAAGAGCTCCTTTACAACCAATTATGCGAGTGAAGCAAAACCCAGTGTGATT 789
Db 121 TCATGTCCTCTCACCTGCGCCCTACGAGAGGCGAGAGATAGCAAGAAGTGGCGCAT 180
|||||
Cp 788 TCATGTCCTCTCACCTGCGCCCTACGAGAGGCGAGAGATAGCAAGAAGTGGCGCAT 729
Db 181 TAAATGTTAGCTGAGGAATTAACAAATACACTGATGATAGCAACAGATTGTGTTG 240
|||||
Cp 728 TAAATGTTAGCTGAGGAATTAACAAATACACTGATGATAGCAACAGATTGTGTTG 669
Db 241 AGACATGCTGTGGGAGAGTCCGCGAGTGTCAAGTCCCTACAGGAGAGGAATCTGAC 300
|||||
Cp 668 AGACATGCTGTGGGAGAGTCCGCGAGTGTCAAGTCCCTACAGGAGAGGAATCTGAC 609
Db 301 TCCGCGCTGAGAGTCTGTGTGAGACATGCTGTACGGGGCTCCACCTGGATGCCATT 360
|||||
Cp 608 TCCGCGCTGAGAGTCTGTGTGAGACATGCTGTACGGGGCTCCACCTGGATGCCATT 549
Db 361 CTGGTGCAG 369
|||||
Cp 548 GGTTGCAG 540

RESULT 3
LOCUS AAB01447 374 bp mRNA EST 30-APR-1998
DEFINITION EST190944 Normalized rat spleen, Bento Soares Rattus sp. cDNA clone
ACCESSION AAB01447
NID 92864402
KEYWORDS EST.
SOURCE Rattus sp.
ORGANISM Rattus sp.
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

TITLE	Waterston, R.
JOURNAL	The WashU-HMNI Mouse EST Project
COMMENT	Unpublished (1996)
CONTACT:	Marria M/Mouse EST Project
WASHU-HMNI MOUSE EST PROJECT	
WASHINGTON UNIVERSITY SCHOOL OF MEDICINE	
4444 FOREST PARK PARKWAY, BOX 8501, ST. LOUIS, MO 63108	
TEL: 314 286 1800	
FAX: 314 286 1810	
EMAIL: mouseest@waterston.wustl.edu	
THIS CLONE IS AVAILABLE ROYALTY-FREE THROUGH LLNL; CONTACT THE	
IMAGE CONSORTIUM (info@image.llnl.gov) FOR FURTHER INFORMATION.	
MG: 228831	
SEQ PRIMER: mob, REGA+ET	
HIGH QUALITY SEQUENCE STOP: 351.	
FEATURES	
SOURCE	
1..362	
/organism="Mus musculus"	
/strain="C57BL/6J"	
/note="Vector: p7T3D-Pac (Pharmacia) with a modified	
polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA	
was primed with a Not I - oligo(dT) primer [5',	
TGTTACCACTCTGAAGTGGAGCGCGCCGGAATTTTTTTTTTTTTTTTTT	
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2	
14.5dpc embryos (total RNA provided by Minoru Ko, Wayne	
State Univ., from 2 1] double-stranded cDNA was ligated to	
Eco RI adaptors (Pharmacia), digested with Not I and	
cloned into the Not I and Eco RI sites of the modified	
p7T3 vector. Library went through one round of	
normalization, and was constructed by Bento Soares and	
M. Fatima Bonaudo."	
/db_xref="taxon:10090"	
/clone="367399"	
/clone_lib="Soares mouse embryo NDMEL3.5 14.5"	
/sex="unknown"	
/tissue_type="embryo"	
/dev_stage="13.5-14.5dpc total fetus"	
/lab_host="DH10B"	
<1..>362	
BASE COUNT	
ORIGIN	
76 a 121 c 89 g 76 t	
Query Match	23.7%; Score 221; DB 20; Length 362;
Best Local Similarity 90.2%; Pred. No. 2,498-227;	
Matches 248; Conservative 0; Mismatches 27; Indels 0; Gaps 0;	
Db 88 CTCCTCAGCGTGTTCACGAGCTGTCCACGAGGAGCTCCTACCTCCACGAGAGAGCTCCAC 147	
QY 1 CTCCTCAGCGTGTTCACGAGCTGTCCACGAGGAGCTCCTACCTCCACGAGAGAGCTCCAC 60	
Db 148 CCAGCTCCAGCAGCTTCAACTAAAGCGTGCATCCCTCCAGTGTGAGAGAGAGGCTTTC 207	
QY 61 CCGGCTCCGGGCGAGCTTCAACTGAAGTGTGCATCTCCAGTGTGAGAGAGAGGCTTTC 120	
Db 208 CCGCGCCCTCTCTGTGAGCTGTATGACACCAAGTCAATGAGCACTGGCTCCGGGCGAGCTCAGC 267	
QY 121 CCGCGCCCTCTCTGTGAGCTGTATGACACCAAGTCAATGAGCACTGGCTCCGGGCGAGCTCAGC 180	
Db 268 CCTGCTGACCCAGAGAGCTGTGTCTACGCTCTTTACCAAGCCAAAGGCTCGAGATGACG 327	
QY 181 CCTGCTGATCCAGAGAGCTACCTCCGCTGCTCTTTACCAAGTGTGAAAGGCTCGAGATGACG 240	
Db 328 CACCTGAAGAGAGATGCGCGGTGTCCGAGGCTTGGT 362	
QY 241 CACCTGAAGAGAGATGCGCGGTGTCCGAGGAGTGTGT 275	
RESULT 5 A1153441 372 bp mRNA EST 30-SEP-1998	
LOCUS uc53c08.r1 Soares 2NDMT Mus musculus cDNA clone 1429358 5', mRNA	
DEFINITION sequence.	
ACCESSION A1153441	

NID	93681910
KEYWORDS	Est.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 372)
AUTHORS	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisell,S., Kubicab,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Thieling,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterson,R.
TITLE	The Mashu-HMI Mouse EST Project
JOURNAL	Unpublished (1996)
COMMENT	Contact: Marra M/Mouse EST Project Mashu-HMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@wustl.wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Meg primer: -28mJ3 rev2 ET from Amersham Seq primer: High quality sequence stop: 193. Location/Qualifiers 1..372 /organism="Mus musculus" /strain="C57BL/6J" /note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - o1190(dry) primer [5' TGTTACCATCTGGAAGTGGAGCGGCCGCCTTTTTTTTTTTTTTTTTT 3'] : double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Bertrand Jordan. Library went through two rounds of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo." /db_xref="taxon:10090" /clone="1429358" /clone_lib="Soares 2ndMT" /sex="male" /tissue_type="Thymus" /dev_stage="4 weeks" /lab_host="DH10B"
FEATURES	
SOURCE	
BASE COUNT	78 a 127 c 81 g 86 t
ORIGIN	
Query Match	20.7%; Score 193; DB 18; Length 372;
Best Local Similarity	94.4%; Pred. N.1.79e-192;
Matches	221; Conservative 0; Mismatches 10; Indels 3; Gaps 3;
Db	7 GTGAGTTATGTAGGGGACTGTGTCCTTAGAGATGACAGTCAAGTCAAGCCGGGACC 66
Oy	425 GTGAGTTATGTAGGGGAGTACTGTGTCCTTAGAGATGAGT-CAGCCAAAGCCGCCGACT 483
Db	67 CTGACACAGAATG-TAATGTAGGCAGAAAGAGGCCCTCCAGCTGACCGGCACTGC 125
Oy	484 CTCGACACGAATAGTAATGTAGGCAGAAAGAGGCCCTCCAGCTGACCGGCACTGC 543
Db	126 AGACCATGAGATCAGGTGATCCCCCTAACAGATGTGTCCACTCCCAAAGCTCGAG 185
Oy	544 A-ACCATTAGATCCAGGTGAGCCGCCCTTACAGCATGTGTCCACACCAAGACTCGAG 602
Db	186 CCGGAGTTCAGATTCCTCCCTCCCGAGGACTGAGCACCCACAGCACTCTCCC 239
Oy	603 CCGGAGTTCAGATTCCTCCCTCCCGAGGACTGAAACACCCGGCGACACTCTCC 656
RESULT	6

LOCUS	249 bp	mrna	EST	21-JAN-1997
DEFINITION	m70d02.r1 Soares mouse placenta 4NDMP13.5 14.5 Mus musculus cDNA			
ACCESSION	AA023504			
NID	91487228			
KEYWORDS	EST.			
SOURCE	house mouse.			
ORGANISM	Mus musculus			
	Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Rodentia; Scleroptera; Muridae; Murinae;			
REFERENCE	Mus.			
AUTHORS	(bases 1 to 249)			
	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucala,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theisberg,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.			
TITLE	The WashU-HMI Mouse EST Project			
JOURNAL	Unpublished (1996)			
COMMENT	Contact: Marra M/Mouse EST Project WashU-HMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@wustl.edu This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information. MGI:273179 Seq primer: -28M13 rev2 from Amersham High quality sequence stop: 196. Location/Qualifiers 1. 249 /organism="Mus musculus" /strain="C57BL/6J" /note-Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dt) primer [5', TGTTACCAATCTGAAGTGGAGCGCGCCGGAATTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo." /db_xref="taxon:10090" /clone="456291" /clone_lib="Soares mouse placenta 4NDMP13.5 14.5" /sex="unknown" /dev_stage="adult" /lab_host="DH10B" <1. >249			
FEATURES	source			
ORIGIN	BASE COUNT 50 a 72 c 78 g 49 t			
	Query Match 18.9%; Score 176; DB 22; Length 249; Best Local Similarity 87.6%; Pred. No. 1,87e-111; Mismatches 29; Indels 0; Gaps 0; Matches 205; Conservative 0;			
Db	15 GCATCTCCAGTGTGAAGAGAGGTCCTCCCGCCCTCTGTGACTGATGCACCAAG 74			
Qy	92 GCATCTCCAGTGTGAAGAGAGGTCCTCCCGCCCTCTGTGACTGATGCACCAAG 151			
Db	75 TCATGCGCAATGTCGTGGGCGAGCTCAACCTGCTGACACCCAGAGCTTGTGACGTGCTC 134			
Qy	152 CCATGGCCAGTACTCTAGACGCTCACCCCTGTGATCCAGAGTCTACGCTCCGTGCTC 211			
Db	135 TTATACCAAGAGCGCTCGAGATGACGACCTGTAAGAGATGCGCGGTGCGGACT 194			
Qy	212 TTATACCAAGAGCGCTCGAGATGACGACCTGTAAGAGATGCGCGGTGCGGACT 271			
Db	195 TGTGCAAGTGCAGAGTGCAGAGCTGCGGCAAGATGCTGAGCTTGGCGCAATG 248			

OY 272 TGTGCAAGCCGAGACGAGCCTGAGCAGATGCGACGCTGTGCAGATG 325
 RESULT 7
 LOCUS AA648997 445 bp mRNA EST 13-NOV-1997
 DEFINITION n541g11.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1186244,
 mRNA sequence.
 ACCESSION AA648997
 NID g2575426
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryote; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
 Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 445)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 unknown library type
 Insert length: 659 Std Error: 0.00
 Seq primer: -40ml3 fwd. ET from Amersham
 High quality sequence stop: 419.
 Location/Qualifiers
 1. 445
 /organism="Homo sapiens"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was prepared from human tonsillar cells enriched for
 germinal center B cells by flow sorting (CD20⁺, 19D-),
 provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
 (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
 primed with a Not I - oligo(dT) primer
 [5'-TGTACCAATCTCAAGTGGAGGCGCCCTCATTTTCTTTTCTTTT-
 3']. Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT73 vector. Library
 went through one round of normalization, and was
 constructed by Bento Soares and M. Fatima Bonaldo."
 /db_xref="taxon:9606"
 /clone_image="1186244"
 /clone_idb="NCI_CGAP_GCB1"
 /tissue_type="germinal center B cell"
 /lab_host="DH10B"
 /lab_host="DH10B"
 BASE COUNT 130 a 82 c 130 g 102 t 1 others
 ORIGIN
 Query Match 17.8%; Score 166; DB 10; Length 445;
 Best Local Similarity 86.5%; Pred. No. 3.73e-159;
 Matches 221; Conservative 0; Mismatches 30; Indels 6; Gaps 5;
 Db 1 TTTTATTTTACCATTT-GCTCAGTAGAAGTTTATGAGAAATCGTTTAAACAA 59
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Cp 915 TTTTATTTTACCATTTGCTCAGTAGAAGTTTATGAGAAATCGTTTAAACAA 858
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 60 TCAGTCAAAAAGACAGCTCTTTTACAAACAAGTTATGGCAGTGGCAAGTAAACCCCA 119
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Cp 857 TCAGTCAAAAAGACAGCTCTTTTACAAACAAGTTATGGCAGTGGCAAGTAAACCCCA 798
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 120 GGTCAATTTCTTATTCCTTACCTGCGCTAGAACGAGNAGCGCGGTGAGCAGAG 179
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Cp 797 GGTCAATTTCTTATTCCTTACCTGCGCTAGAACGAGNAGCGCGGTGAGCAGAG 741
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 180 AGATGGGCTTATGGAATGTAAGTAGAGAAATTCAAAAATACACTGTGATGACAA 239
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Cp 740 AGATGGGCTTATGGAATGTAAGTAGAGAAATTCAAAAATACACTGTGATGACAA 681
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 240 AGGTTGTGTGAACATGCGAGGGG 266
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Cp 680 AGATTTGTGTGAGACATGCTGTGGG 654
 RESULT 8
 LOCUS AA845221 319 bp mRNA EST 04-MAR-1998
 DEFINITION BK76612.s1 Barstead spleen HPLRB2 Homo sapiens cDNA clone
 IMAGE:1413815 3', mRNA sequence.
 ACCESSION AA845221
 NID g2931672
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryote; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
 Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 319)
 AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Gelsel, G., Jost, S.,
 Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,
 Martin, J., Moore, B., Schellenberg, R., Steptoe, M., Tan, F.,
 Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.
 WashU-NCI human EST Project
 JOURNAL Unpublished (1997)
 COMMENT
 Contact: Wilson, R.K.
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Seq primer: -40ml3 fwd. ET from Amersham
 High quality sequence stop: 189.
 Location/Qualifiers
 1. 319
 /organism="Homo sapiens"
 /note="Organ: spleen; Vector: pT73D-Pac (Pharmacia) with
 a modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st
 strand cDNA was primed with a Not I - oligo(dT) primer [5'
 TGTACCAATCTCAAGTGGAGGCGCCCTCATTTTCTTTTCTTTT-
 3']; double-stranded cDNA was ligated to Eco RI adaptors
 [AATTCGATCCTTG], digested with Not I and cloned into the
 Not I and Eco RI sites of the modified pT73 vector.
 Library constructed by Bob Barstead."
 /db_xref="taxon:9606"
 /clone_image="1413815"
 /clone_idb="Barstead spleen HPLRB2"
 /tissue_type="spleen"
 /sex="male"
 /dev_stage="adult, 17 years"
 /lab_host="DH10B"
 /lab_host="DH10B"
 BASE COUNT 102 a 56 c 87 g 74 t
 ORIGIN
 Query Match 17.4%; Score 162; DB 13; Length 319;
 Best Local Similarity 86.1%; Pred. No. 3.00e-154;
 Matches 229; Conservative 0; Mismatches 31; Indels 6; Gaps 5;
 Db 1 TTTTATTTTACCATTT-GCTCAGTAGAAGTTTATGAGAAATCGTTTAAACAA 59
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Cp 914 TTTTATTTTACCATTTTGTCTCAGTAGAAGTTTATGAGAAATCGTTTAAACAA 857
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 60 CAGTCAAAAAGACAGCTCTTTTACAAACAAGTTATGGCAGTGGCAAGTAAACCCCA 119
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Cp 856 CAGTCAAAAAGACAGCTCTTTTACAAACAAGTTATGGCAGTGGCAAGTAAACCCCA 797
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 120 GTCAATTTCTTATTCCTTACCTGCGCTAGAACGAGNAGCGCGGTGAGCAGAG 179
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Cp 796 GTCTGATTTCTATGTTCCCTTACCTGCGCTAGAACGAGNAGCGCGGTGAGCAGAG 740
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 180 GATGGGCTTATGGAATGTAAGTAGAGAAATTCAAAAATACACTGTGATGACAA 239
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Cp 739 GATGGGCTTATGGAATGTAAGTAGAGAAATTCAAAAATACACTGTGATGACAA 680
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 240 GGGTTGTGTGAACATGCGAGGGG 265
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Cp 679 GGATGTGTCAGACATGCTGTGGG 654

RESULT 9 AA760677 365 bp mRNA EST 18-FEB-1998
LOCUS n209c12.s1 NCI_CGAP_GCB1 Homo sapiens CDNA clone IMAGE:1287286,
DEFINITION mRNA sequence.
ACCESSION AA760677
NID 92809607
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 365)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bdrr/image/image.html

Insert Length: 992 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 363.
Location/Qualifiers
FEATURES
source

1..365
/organism="Homo sapiens"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site.1: Not I; Site.2: Eco RI; 1st strand CDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). CDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGAAGTGGAGCGCGCCGCTCATTTTCTTTT-
3']. Double-stranded CDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
/db_xref="taxon:9606"
/clone_lib="IMAGE:1287286"
/clone_lib="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
BASE COUNT 108 a 67 c 104 g 86 t
ORIGIN

Query Match 16.6%; Score 155; DB 13; Length 365;
Best Local Similarity 85.7%; Pred. No. 1.10e-145;
Matches 216; Conservative 0; Mismatches 31; Indels 5; Gaps 4;
Db 10 TTTTGCTAGTAGAGTTTAAATGAGAAATCGTTGTTTAAACATCAGTCAAAAGAAC 69
|||||
Cp 900 TTTTGCTAGTAGAGTTTAAATGAGAAATCGTTGTTTAAACATCAGTCAAAAGAAC 843
|||||
Db 70 AGCTCTTTTCAAAACAGTTATGCGAAGTCAAAACCCAGGTCATTTCTAT 129
|||||
Cp 842 AGCTCTTTTCAAAACAGTTATGCGAAGTCAAAACCCAGGTCATTTCTAT 783
|||||

Db 130 TCCTTACCTGCCCCCTAGAGAGCGGCTGACGAGAGATGGGCTATTGA 189
|||||
Cp 782 TCCTTACCTGCCCCCTAGAGAGCGGCTGACGAGAGAGC -AGA -GGATGAGCAAGA -AGATGGGCGCATTTAA 726
|||||
Db 190 AATGGTAGCTAGAGGAATTAACAAATACACTGATGTAGCAACAGGGTGTGGTGA 249
|||||
Cp 725 AATGTATGCTGGAGGAATTAACAAATACACTGATGTAGCAACAGGATTTGTGTAGA 666
|||||
Db 250 CATGCCAGGGGG 261
|||||
Cp 665 CATGCTGTGGGG 654
|||||

RESULT 10 AA283020 397 bp mRNA EST 14-AUG-1997
LOCUS z16b04.s1 NCI_CGAP_GCB1 Homo sapiens CDNA clone IMAGE:713263 3',
DEFINITION mRNA sequence.
ACCESSION AA283020
NID 91925944
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Eukaryota; Metazoa; Chordata;
Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae;
Homo.

REFERENCE 1 (bases 1 to 397)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 685 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham.

FEATURES
source
1..397
/organism="Homo sapiens"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site.1: Not I; Site.2: Eco RI; 1st strand CDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). CDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGAAGTGGAGCGCGCCGCTCATTTTCTTTT-
3']. Double-stranded CDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
/db_xref="taxon:9606"
/clone_lib="IMAGE:713263"
/clone_lib="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
BASE COUNT 115 a 76 c 121 g 85 t
ORIGIN

Query Match 16.5%; Score 154; DB 8; Length 397;
Best Local Similarity 85.7%; Pred. No. 1.83e-144;
Matches 215; Conservative 0; Mismatches 31; Indels 5; Gaps 4;
Db 6 TTTTGCTAGTAGAGTTTAAATGAGAAATCGTTGTTTAAACATCAGTCAAAAGAAC 65
|||||
Cp 899 TTTTGCTAGTAGAGTTTAAATGAGAAATCGTTGTTTAAACATCAGTCAAAAGAAC 842
|||||
Db 66 GCTCTTTTCAAAACAGTTATGCGAAGTCAAAACCCAGGTCATTTCTAT 125
|||||

Cp	841	GCTCCTTTACAAACCAGTATGCGAGTGGTGAGGGCAAAAACCCAGGTCTGATTTCATGTT	782
Db	126	CCTTTCACCTGCCCTCTAGAAGCCGAAGAGCGCGGTGCAGCAGAGCATGGGCTTTGAA	185
Cp	791	CTCTCACCTTGCCTCTACGAGAGGGCC-AGA-GGATGAGCACA-A-GATGGGGCCATTAA	725
Db	186	ATGAGAGCTAGAGCAATTACAAAAATPACACTGTGATGTAGCAACAGGGTTGGTGAAC	245
Cp	724	ATGTTAGCTGAGGAAATTACAAAAATPACACTGTGATGTAGCAACAGGATTTGTGTAGAC	665
Db	246	ATGCCAGGGGG	256
Cp	664	ATGCTGTGGGG	654

RESULT	11
LOCUS	AA282757
DEFINITION	2891807.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:104820 3'
ACCESSION	AA282757
NID	g1825691
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
AUTHORS	Vertebrates; Mammalia; Eutheria; Primates; Catarrhini; Homnidae;
TITLE	Homo.
JOURNAL	1 (bases 1 to 431)
COMMENT	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997) Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert.Strausberg@nih.gov This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Insert Length: 1027 Std Error: 0.00 Seq primer: -41m13 fwd. ET from Amersham High quality sequence stop: 400. Location/Qualifiers 1. 431 /organism="Homo sapiens" /note="Vector: pYT3D-Pac (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from human tonsillar cells enriched for germlinal center B cells by flow sorting (CD20+ IgD-), provided by Dr. Louis M. Staudt (NCI), Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - oligo(dT) primer [5'-GTGACCAATCTGACGAGGAGCGCGCCCTCATTTTTTTTTTTTTT- 3']". Double-stranded cDNA was ligated to Eco RI adaptors and Eco RI sites of the modified pYT3D vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo." /db_xref="taxon:9606" /clone="IMAGE:704820" /clone_lib="NCI CGAP GCB1" /tissue_type="germlinal center B cell" /lab_host="DH10B" complement(<1. >431) /db_xref="GDB:5854544"

BASE COUNT	122 a	85 c	131 g	93 t
ORIGIN				

Query Match	16.5%	Score 154:	DB 8;	Length 431:
Best Local Similarity	85.7%	Pred. No. 1,83e-144;		
Matches	215;	Conservative	0;	Mismatches 31;
			Indels	5;
			Gaps	4;

6	TTTGTCAGTAGAAGTTTAATGAGAAACGCTGTTTAAAGCAATCAGCAAAAGAAC	65

CP	899	TTTGCTGAGTAGAGTTAATGAGAA--GTGTTTCAACATCACTGCAAAAAGACA	842
-Db	66	GCTCTTTTACAAACAGTTATGCGAGTGCAGATCAAAACCCAGTTCATTTCTATT	125
CP	841	.GCTCTTTTACAAACAGTTATGCGAGTGCAGATGCAGCAAAACCCAGTTCATTTCTATT	782
Db	126	CCTTTCACCTGCCCCCTAGAAAGCCGAAGGCGGGGTGAGCAGAGAGATGGGGCTATTGAA	185
CP	781	CCTTCACCTGCCCCCTAGAGAGGGC-AGA-GGATGACCAAGA-AGATGGGGCCATTGAA	725
Db	186	ATGGTAGTAGAGCAATATACAAATATACACTCTGTGTGTAGCAACAGGGTTGGTGAAC	245
CP	724	ATGTAGCTGAGAGCAATTACAAAATACACTCTGTGTGTAGCAACAGGATTTGGTGAAC	665
Db	246	ATGCCAGGGGG 256	
CP	664	ATGCTGTGGGG 654	
RESULT	12		
LOCUS	AA806350	431 bp	mRNA
DEFINITION	oc6cf09.s1 NCI_CGAP_GCB1 Homo sapiens	CDNA clone	IMAGE:1350857 3'
ACCESSION	AA806350		
NID	g2875100		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryotes: Metazoa: Chordata: Vertebrata; Mammalia; Eutheria;		
AUTHORS	Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 431)		
JOURNAL	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .		
COMMENT	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index		
	Unpublished (1997)		
	Contact: Robert Strausberg, Ph.D.		
	Tel: (301) 496-1550		
	Email: Robert.Strausberg@nih.gov		
	Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,		
	Ph.D., Gerald Marti, M.D.		
	CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima		
	Bonaldo, Ph.D.		
	CDNA Library Arrayed by: Greg Lennon, Ph.D.		
	DNA sequencing by: Washington University Genome Sequencing Center		
	clone distribution: NCI-CGAP clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LNLT at:		
	www-bio.1lnl.gov/bdnp/image/image.html		
	Insert Length: 1103 Std Error: 0.00		
	Sed primer: -40ml3 fwd. Ef from Amersham.		
FEATURES	Location/Qualifiers		
SOURCE	1. 431		
	/organism="Homo sapiens"		
	/note="Vector: pRT73D-Pac (Pharmacia) with a modified		
	/polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand CDNA		
	was prepared from human tonsillar cells enriched for		
	geminal center B cells by flow sorting (CD20+, IgD-),		
	provided by Dr. Louis M. Staudt (NCI), Dr. David Allman		
	(NCI) and Dr. Gerald Marti (CEBR). CDNA synthesis was		
	primed with a Not I - oligo(dT) primer		
	[5'-GTGTACCAATCTGAGTGGAGGCGGCTCATTTTTTTTTTTTTTTTTT-		
	3']. Double-stranded CDNA was ligated to Eco RI adaptors		
	(Pharmacia), digested with Not I and cloned into the Not I		
	and Eco RI sites of the modified pRT73 vector. Library		
	went through one round of normalization, and was		
	constructed by Bento Soares and M. Fatima Bonaldo."		
	/db_xref="taxon:9606"		
	/clone="IMAGE:1350857"		
	/clone_11b="NCI CGAP GCB1"		
	/tissue_type="germinal center B cell"		
	/lab_host="DH10B"		
BASE COUNT	125 a 80 c 133 g 93 t		


```

/db_xref="taxon:9606"
/clone_lib="human fetal brain"
/dev_stage="fetus"
/tissue_type="brain"
BASE COUNT      94 a      54 c      83 g      64 t      10 others
ORIGIN

Query Match      15.7%; Score 146; DB 7; Length 305;
Best Local Similarity 84.2%; Pred. No. 1,01e-134;
Matches 213; Conservative 4; Mismatches 29; Indels 7; Gaps 3;

Dh 6 TTTGCTCAGTAGAATTTTAATGAGAGAAATCGTTGTTTAAACAAATCAGTCAGCAAAAAAGACA 65
    |||
Cp 899 TTGCTCAGTAGAATTTTAATGAGAGAAA--GTTGTTTCAAAATCAGTCAGCAAAAAAGACA 842
    |||

Dh 66 GCTCTTTTCAAAACAGATTGACAGTGGCAAGTCAAAAACCCAGGTTCAATTTTCTATT 125
    |||
Cp 841 GCTCTTTTCAAAACAGATTATGACAGTGGTGAAGGCAAAACCCAGGTTCAATTTTCTATT 782
    |||

Dh 126 CCTTTCACCCSCCCTAGAGAGGGGCAAGAGSCGGGTGACAGAGAGATGGGGCTATTG 185
    |||
Cp 781 CCTCTACCTGGCCCTCAGAGAGAGGGCAGAGGA---TGAGCAAGA-AGATGGGGCCATT 727
    |||

Dh 186 AAATGCTAGCTTAGAGGAATTTACAAAAATTAACCTCTGABGTAGCAACAGGGTTGTGCTGA 245
    |||
Cp 726 AAATTTAGCTTGAGGAATTTACAAAAATTAACCTCTGATGTATACCAACAGGATTTGTGTGAG 667
    |||

Dh 246 ACATGCCAGGGGG 258
    |||
Cp 666 ACATGCTGTGGGG 654

```

[illegible]

The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13 Reverse
Location/Qualifiers
1. 499

```

/organism="Rattus sp."
/ntex="Organ: spleen; Vector: pF773Pac; Site_1: EcoRI
Site_2: NotI"
/db_xref="ATCC (inhost):2009517"
/db_xref="taxon:10118"
/clone="NSPAA76"
/clone_1b="Normalized rat spleen, Bento Soares"
BASE COUNT      93 a      155 c      165 g      84 t      2 others
ORIGIN

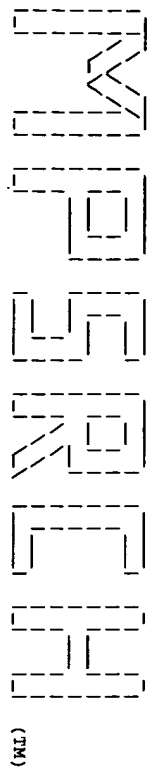
```

Query Match	14.28;	Score 132;	DB 14;	Length 499;
Best Local Similarity	97.88;	Pred. No. 8.90e-118;		
Matches 134;	Conservative 0;	Mismatches 3;	Indels 0;	Gaps 0;

Db 363 GTAAGCATGTGTGCACACCAAGACCTGCAGCCGGAGTCAGATTCCTCCTCCCTGA 422
 ||| |||||
 QY 571 GTACAGCATGTGTGCACACCAAGACCTGCAGCCGGAGTCAGATTCCTCCTCCCTGA 630
 ||| |||||
 Db 423 GGCACGTGAACACCGCGGCACCTCCCGACAGCATGTCTCACCAATCTGTGCTACAT 482
 ||| |||||
 QY 631 GGCACGTGAACACCGCGGCACCTCCCGACAGCATGTCTCACCAATCTGTGCTACAT 690
 ||| |||||
 Db 483 CAGAGTGTATTTTGTGA 499
 ||| |||||
 QY 691 CAGAGTGTATTTTGTGA 707

Search completed: Tue Apr 27 11:26:55 1999
Job time : 1354 secs.

THIS PAGE BLANK (USPTO)



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MSearch_tpn n.a. - n.a. Smith-Waterman search, using a protein query
which has been backtranslated into n.a. using IUPAC symbols

Run on: Tue Apr 27 10:31:00 1999; MasPar time 99.06 Seconds
Tabular output not generated. 572.525 Million cell updates/sec

Title: >US-09-011-797-4
Description: (1-8) from US09011797.pep
Perfect Score: 100
N.A. Sequence: 1 ACNNTNCAYCARAYGNAAYGTN 24
Comp: TGNRANGTGTTCRCNTTCAN

Scoring table: TABLE bcktranslated
Gap 40

Mmatch STD : Dbase 0; Query 0

Searched: 602357 seqs, 1181590623 bases x 2

Post-processing: Minimum Match 08
Listing first 45 summaries

Database: emb157
1:em_ba 2:em_fun 3:em_htg 4:em_hum1 5:em_hum2 6:em_in
7:em_om 8:em_or 9:em_ov 10:em_pat 11:em_ph 12:em_pl
13:em_ro 14:em_un 15:em_v1
Database: genbank110
16:gb_pat 17:gb_ba2 18:gb_htg 19:gb_in 20:gb_om 21:gb_ov
22:gb_pat 23:gb_ph 24:gb_pl1 25:gb_pl2 26:gb_pl1
27:gb_pr2 28:gb_pr3 29:gb_ro 30:gb_st 31:gb_sts 32:gb_sy
33:gb_un 34:gb_v1

Statistics: Mean 39.423; Variance 98.948; scale 0.398
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	100	100.0	372	29	MMU44027 Mus musculus orphanin	5.45e+00
2	100	100.0	633	29	MMPPNMRNA M.musculus mRNA for pr	5.45e+00
3	100	100.0	878	26	HSPNPNEX3 H.sepiens gene encodin	5.45e+00
4	100	100.0	925	29	579730 OR1L receptor agonist	5.45e+00
5	100	100.0	972	29	RNU48262 Rattus norvegicus pre-	5.45e+00
6	100	100.0	988	29	RNPNNMRNA R.norvegicus mRNA for	5.45e+00
7	100	100.0	1015	26	HSPNPNMRNA H.sepiens mRNA for pre	5.45e+00
8	100	100.0	1047	29	MUSN23K House N23K mRNA for N2	5.45e+00
9	100	100.0	1198	28	HSU48263 Human pre-pro-orphanin	5.45e+00
10	100	100.0	1253	29	MMPPNEX3 M.musculus gene encodi	5.45e+00
11	100	100.0	1354	29	MUSNOP Mouse mRNA for nocicep	5.45e+00
12	95	95.0	1209	20	AB005251 Bos taurus mRNA for no	1.93e+01

13	90	90.0	2258	24	SCYBI020W S.cerevisiae chromosome	6.60e+01
14	90	90.0	2551	24	SCU15087 Saccharomyces cerevisi	6.60e+01
15	90	90.0	3061	24	YSCHAP S.cerevisiae HAP3 locu	6.60e+01
16	90	90.0	150024	18	HS5222P13 Human DNA sequence ***	6.60e+01
17	90	90.0	199092	18	HS80N2 Human DNA sequence ***	6.60e+01
18	85	85.0	228	19	ECU03705 Elaphidium clavatum inc	2.16e+02
19	85	85.0	248	29	RRMACPE05 Rattus norvegicus Wist	2.16e+02
20	85	85.0	1080	29	RATPHOCAMB Rat CAMP phosphodiester	2.16e+02
21	85	85.0	1080	29	RATPDE4A Rat ratPDE4 mRNA encod	2.16e+02
22	85	85.0	1086	25	AF071891 Prunus armeniaca 40S r	2.16e+02
23	85	85.0	1158	27	AB006077 Homo sapiens doc-1 mRN	2.16e+02
24	85	85.0	1193	24	LEGPF46 L.esculentum GBFA mRNA	2.16e+02
25	85	85.0	1704	24	SCU36503 Saccharomyces cerevisi	2.16e+02
26	85	85.0	1888	24	SCMRRI130 Saccharomyces cerevisi	2.16e+02
27	85	85.0	1888	24	SCMRRI130 Saccharomyces cerevisi	2.16e+02
28	85	85.0	2036	19	AB006675 Halocynthia roretzi Hr	2.16e+02
29	85	85.0	2158	22	122474 Sequence 3 from patent	2.16e+02
30	85	85.0	2158	29	RATDPD Rat CAMP phosphodiester	2.16e+02
31	85	85.0	2254	21	XIASURAB Xenopus laevis mRNA fo	2.16e+02
32	85	85.0	2263	21	AF019906 Xenopus laevis transme	2.16e+02
33	85	85.0	2268	24	SCYDL127W S.cerevisiae chromosome	2.16e+02
34	85	85.0	2647	29	RATPHOB8 Rattus norvegicus phos	2.16e+02
35	85	85.0	2814	19	SLMARDD Physarum polycephalum	2.16e+02
36	85	85.0	3133	29	RNU95748 Rattus norvegicus CAMP	2.16e+02
37	85	85.0	7030	17	SMASSPH Serratia marcescens DN	2.16e+02
38	85	85.0	10510	16	U32761 Haemophilus influenzae	2.16e+02
39	85	85.0	12882	16	AEO00911 Methanobacterium therm	2.16e+02
40	85	85.0	42729	16	MECY71 Mycobacterium tubercu	2.16e+02
41	85	85.0	75317	24	SC4357 S.cerevisiae chromosome	2.16e+02
42	85	85.0	103319	27	AC004217 Homo sapiens 12q24.1 p	2.16e+02
43	85	85.0	110000	24	SCCHRXY S.cerevisiae chromoso	2.16e+02
44	85	85.0	162575	18	AC004086 *** SEQUENCING IN PRO	2.16e+02
45	85	85.0	170030	28	AC005829 Homo sapiens Chromosom	2.16e+02

ALIGNMENTS

RESULT 1	LOCUS	MMU44027	372 bp	mRNA	ROD	30-MAY-1996
	DEFINITION	Mus musculus orphanin FQ/nociceptin precursor mRNA, partial cds.				
	ACCESSION	U44027				
	NID	g1335869				
	KEYWORDS	house mouse.				
	SOURCE	Mus musculus.				
	ORGANISM	Eukaryota; Eutherochordal eukaryotes; Metazoa; Chordata; Vertebrata; Mithidia; Rodentia; Sciurognathi; Myomorpha; Muridae; Murinae; Mus.				
	REFERENCE	1 (bases 1 to 372)				
	AUTHORS	Pan,Y.X., Xu,J. and Pasternak,G.W.				
	TITLE	Cloning and expression of a cDNA encoding a mouse brain orphanin FQ/nociceptin precursor				
	JOURNAL	Biochem. J. 315 (Pt 1), 11-13 (1996)				
	REFERENCE	96207555				
	AUTHORS	2 (bases 1 to 372)				
	TITLE	Direct Submission				
	JOURNAL	Submitted (28-DEC-1995) Ying-Xian Pan, Neurology, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10021, USA				
	FEATURES	location/Qualifiers				
	source	1..372				
		/organism="Mus musculus"				
		/db_xref="taxon:10090"				
		/tissue_type="Brain"				
		<2..331				
		/codon_start=1				
		/product="orphanin FQ/nociceptin precursor"				
		/db_xref="PID:g1335870"				
		/translation="TTSALYGPKASEQHKLMPRVSLVQVRDAEPGADAEFGADAEPGADAEVQKQKRFQFTGARKSARLANQKRSEFMRYLVMSQSORRT LHKONNV"				
		misc_feature				
		188..238				

GenBank staff at the National Library of Medicine created this entry [NCBI g1bbbsq 171655] from the original journal article.

This sequence comes from Fig. 4.

```
FEATURES
  source      Location/Qualifiers
            1..925
              /organism="Rattus sp."
              /db_xref="taxon:10118"
              1..507
                /partial
                /gene="ORL1 receptor agonist precursor"
                1..507
                  /partial
                  /gene="ORL1 receptor agonist precursor"
                  /note="endogenous agonist of opioid receptor-like ORL1
                    receptor. Authors indicate ORL1 receptor agonist peptide
                    is encoded by residues 122-138. Method: conceptual
                    translation with partial peptide sequencing. This sequence
                    comes from Fig. 4"
                  /codon_start=1
                  /product="ORL1 receptor agonist precursor"
                  /db_xref="PID:g117239"
                  /translation="LSVFSFSSCPEDLTCOERLHPAPGSFNLKLTICCEKYPFPPRL
                    WILCTKAMASDSQSLSPADELTSAALYQSKASEMOKLTKMPVRSVYQARDEPND
                    AEPVADAEADAEVPADEADAEVQKOLKRRGFTGARKSARKLANQRSEFMKRYLVSMOSSORRR
                    TLHQNGNV"
BASE COUNT      222 a      271 c      237 g      195 t
ORIGIN
Query Match      100.0%; Score 100; DB 29; Length 925;
Best Local Similarity 65.2%; Pred. No. 5.45e+00;
Matches 15; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Db 481 ACTTCGACCGAATGTAATGT 503
      11:1111111111111111
      1 ACNTNCAYCARAYGGAAYGT 23

RESULT 5
LOCUS      RNU48262      972 bp      mRNA      ROD      23-AUG-1996
DEFINITION Rattus norvegicus pre-pro-orphanin FQ mRNA, complete cds.
ACCESSION      U48262
NID      g1185011
KEYWORDS
SOURCE      Norway rat.
ORGANISM      Rattus norvegicus
              Eukaryota; Eukaryota; Eukaryota; Metazoa; Chordata;
              Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
              Murinae; Rattus;
              1 (bases 1 to 972)
              Notchaker, H.P., Reinscheid, R.K., Mansour, A., Henningsen, R.A.,
              Ardati, A., Monama, F.J., Jr., Watson, S.J. and Clivelli, O.
              Primary structure and tissue distribution of the orphanin FQ
              precursor
              Proc. Natl. Acad. Sci. U.S.A. 93 (16), 8677-8682 (1996)
JOURNAL      MEDLINE
REFERENCE      2 (bases 1 to 972)
AUTHORS      Notchaker, H.-P. and Henningsen, R.A.
TITLE      Direct Submision
JOURNAL      Submitted (02-FEB-1996) Hans-Peter Notchaker, PRPN 69/202,
              Hoffmann-La Roche AG, Grenzachstr 124, Basel, 4070, Switzerland
FEATURES
  source      Location/Qualifiers
            1..972
              /organism="Rattus norvegicus"
              /strain="Sprague-Dawley"
              /db_xref="taxon:10116"
              28..84
                /codon_start=1
                /product="pre-pro-orphanin FQ"
                /db_xref="PID:g1185012"
                /translation="MKILFCDVLLSLSSVFSFSSCPEDLTCOERLHPAPGSFNLKLC
                  ILQCEKYPFPPRLWILCTKAMASDSQSLSPADELTSAALYQSKASEMOKLTKMPVRSV
                  YQARDEPND AEPVADAEADAEVPADEADAEVQKOLKRRGFTGARKSARKLANQRSEFMK
                  YLVSMOSSORRRTLHQNGNV"
                85..570
                  mat_peptide
                    85..570
```

```
3'UTR      574..972
  polyA_site
BASE COUNT      214 a      293 c      251 g      214 t
ORIGIN
Query Match      100.0%; Score 100; DB 29; Length 972;
Best Local Similarity 65.2%; Pred. No. 5.45e+00;
Matches 15; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Db 547 ACTTCGACCGAATGTAATGT 569
      11:1111111111111111
      1 ACNTNCAYCARAYGGAAYGT 23

RESULT 6
LOCUS      RNP48262      988 bp      RNA      ROD      10-SEP-1996
DEFINITION R norvegicus mRNA for prepronociceptin.
ACCESSION      X97375
NID      g1532064
KEYWORDS      prepronociceptin.
SOURCE      Norway rat.
ORGANISM      Rattus norvegicus
              Eukaryota; Eukaryota; Eukaryota; Metazoa; Chordata;
              Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
              Rattus;
              1 (bases 1 to 988)
              Moller, J.C., Simons, M.J., Soularue, P., Liners, F., Vassart, G.,
              Meunier, J.C. and Parmentier, M.
              Structure, tissue distribution, and chromosomal localization of the
              prepronociceptin gene
              Proc. Natl. Acad. Sci. U.S.A. 93 (16), 8666-8670 (1996)
JOURNAL      MEDLINE
REFERENCE      2 (bases 1 to 988)
AUTHORS      Parmentier, M.
TITLE      Direct Submision
JOURNAL      Submitted (19-APR-1996) M. Parmentier, Universite Libre de
              Bruxelles, 1 R I B N ULB Campus Erasme, 808 Route de Lennik, B-
              1070 Bruxelles, BELGIUM
COMMENT      Related sequence S79730.
FEATURES
  source      Location/Qualifiers
            1..988
              /organism="Rattus norvegicus"
              /db_xref="taxon:10116"
              /clone_1lb="lambda DASHII (Stratagene #943501)"
              /clone_1lb="lambda ZAPII (brain cDNA)"
              /clone="RGE109"
              /clone="RB101"
              <1..150
                /number=2
                /join(<1..150,151..615,616..974)
                25..69
                  join(25..150,151..570)
                  /codon_start=1
                  /product="prepronociceptin"
                  /db_xref="PID:e244795"
                  /db_xref="PID:g1532065"
                  /translation="MKILFCDVLLSLSSVFSFSSCPEDLTCOERLHPAPGSFNLKLC
                    ILQCEKYPFPPRLWILCTKAMASDSQSLSPADELTSAALYQSKASEMOKLTKMPVRSV
                    YQARDEPND AEPVADAEADAEVPADEADAEVQKOLKRRGFTGARKSARKLANQRSEFMK
                    YLVSMOSSORRRTLHQNGNV"
                    70..567
                      /product="prepronociceptin"
                      151..615
                        /number=3
                        exon
                          616..974
                            /number=4
                            exon
                              polyA_site
                                971
                                  BASE COUNT      231 a      292 c      250 g      215 t
                                  ORIGIN
                                  Query Match      100.0%; Score 100; DB 29; Length 988;
```



```

exon
/stratn="129/Sv"
/db_xref="taxon:10090"
/clone_id="lambda fixii (Stratagene #946305)"
/clone="MG43"
605..1087
/number=3
/label=ex3
/usedIn=X87371:MPPN_CDS
/usedIn=X97371:MPPN_mRNA
/product="prepronociceptin"
BASE COUNT      356 a      307 c      323 g      267 t
ORIGIN
Query Match      100.0%; Score 100; DB 29; Length 1253;
Best Local Similarity 65.2%; Pred. No. 5,456+00;
Matches 15; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Db 1016 ACCCGCACGAGATGCTATGT 1038
||:||||:||||:||||:|
Oy 1 ACNTNCAYCARAAVGGNAAYGT 23

RESULT 11
LOCUS MUSNOP 1354 bp mRNA ROD 09-APR-1997
DEFINITION Mouse mRNA for nociceptin/orphanin FQ, complete cds.
ACCESSION D82866
G1311472
KEYWORDS nociceptin/orphanin FQ.
SOURCE Mus musculus adult brain cDNA to mRNA.
ORGANISM Mus musculus
Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae;
Murinae; Mus.
1 (bases 1 to 1354)
Takeshima,H.
Direct Submission
Submitted (25-DEC-1995) to the DDBJ/EMBL/GenBank databases. Hitoshit
Takeshima, University of Tokyo, Dept. of Pharmacology, Faculty of
Medicine; Hongo 7-3-1, Bunkyo-ku, Tokyo 113, Japan
(E-mail:takeshimem-u-tokyo.ac.jp, Tel:03-3812-2111(ex.3422),
Fax:03-3815-9360)
2 (bases 1 to 1354)
Takeshima,H.
Unpublished (1996)
3 (sites)
Houtani,T., Nishi,M., Takeshima,H., Nukada,T. and Sugimoto,T.
Structure and regional distribution of nociceptin/orphanin FQ
precursor
Biochem. Biophys. Res. Commun. 219 (3), 714-719 (1996)
66216718
JOURNAL MEDLINE
FEATURES
SOURCE
1..1354
Location/Qualifiers
/organism="Mus musculus"
/db_xref="taxon:10090"
/dev_stage="adult"
/tissue_type="brain"
360..923
/codon_start=1
/product="nociceptin/orphanin FQ"
/db_xref="PID:d101281"
/db_xref="PID:g1311473"
/translaton="MKILFCDVLLLSLSSVSSCPDCLTCOEKLPAPDSFNLKTC
ILQEEKVPRLPMTVCTKMGASGSLSPADPLVSAALYOPKASMOHLKRPVRYR
SLYVRAEAGADAEFGADAEPCGDDAEVEGOKLORRFGFTCARSAKRLAKQKRF
SEFRKQVLSMSSQRRRLHONGNV"
1317..1322
polya_signal
BASE COUNT      309 a      370 c      386 g      289 t
ORIGIN
Query Match      100.0%; Score 100; DB 29; Length 1354;
Best Local Similarity 65.2%; Pred. No. 5,456+00;
Matches 15; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

```


Db	897	ACCCGTGACCAAGATGCTATGT	919	
Qy	1	ACNTTNCATCARAAATGGAATG	23	
RESULT	12			
LOCUS	AB005251	1209 bp	MM	09-APR-1998
DEFINITION	Bos taurus mRNA for nociceptin/orphanin FQ precursor, complete cds			
ACCESSION	AB005251			
NID	93041766			
KEYWORDS	nociceptin/orphanin FQ precursor.			
SOURCE	Bos taurus cDNA to mRNA.			
ORGANISM	Bos taurus			
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Artiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.			
AUTHORS	1 (sites)			
TITLE	Okuda-Ashtaka, E., Minami, T., Tachibana, S., Yoshinara, Y., Nishinuchi, Y., Kimura, T. and Ito, S.			
JOURNAL	Nucleic Acids Res.			
MEDLINE	1998;126(1):1-11			
REFERENCE	Transmission			
AUTHORS	Nature 392 (6673), 286-289 (1998)			
TITLE	2 (bases 1 to 1209)			
JOURNAL	Okuda-Ashtaka, E.			
MEDLINE	Submitted (26-JUN-1997) to the DDBJ/EMBL/GenBank databases.			
REFERENCE	Okuda-Ashtaka, E., Minami, T., Tachibana, S., Yoshinara, Y., Nishinuchi, Y., Kimura, T. and Ito, S.			
AUTHORS	Chemistry: 10-15 Fumizono, Moriyuchi, Osaka, JAPAN, Moriyuchi, Osaka 570, Japan (E-mail: ashtaka@kai.kmu.ac.jp, tel: 06-992-1001, Fax: 06-992-1781)			
FEATURES	Location/Qualifiers			
SOURCE	1..1209			
CDS	/organism="Bos taurus"			
	/db_xref="taxon:9913"			
	210..740			
	/codon_start=1			
	/product="nociceptin/orphanin FQ precursor"			
	/db_xref="PID:01026347"			
	/db_xref="PID:93041769"			
	/translation="MKILFCDLLILSFSSVSSCKDCLYCREKRLPTLDSFLSNC ILDEEKAFSPMLPTGTVKVARSMQSPADPDHVAALDOPASMOHKKMPRR SLFQKQKTEPLEGVEGIEKQLOKRGRTGAKRSARKLANKRSEFRRQTLVLS MOSSORRTLLDNQNA"			
BASE COUNT	240 a 359 c 341 g 269 t			
ORIGIN				
Query Match	95.0%; Score 95; DB 20; Length 1209;			
Best Local Similarity	63.6%; Pred. No. 1.93e+01;			
Matches	14; Conservative 5; Mismatches 3; Indels 0; Gaps 0;			
Db	714	ACCCGACACCAAGATGCTATG	735	
Qy	1	ACNTTNCATCARAAATGGAATG	22	
RESULT	13			
LOCUS	SCYB1020W	2258 bp	DNA	PLN 11-AUG-1997
DEFINITION	S cerevisiae chromosome II reading frame ORF YBL020W.			
ACCESSION	Z35781	Y13134		
NID	9356014			
KEYWORDS	baker's yeast.			
SOURCE	Saccharomyces cerevisiae			
ORGANISM	Eukaryote; mitochondrial eukaryotes; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.			
REFERENCE	1 (bases 1 to 2258)			
AUTHORS	Goffeau, A., Joniaux, J.-L., Purnelle, B., Skala, J., de Wergifosse, P. and van Dycck, L.			
JOURNAL	Unpublished			
REFERENCE	2 (bases 1 to 2258)			
AUTHORS	MPIS.			
TITLE	Direct Submission			

REFERENCE	AUTHORS	JOURNAL
3 (bases 1 to 2258)	Feldmann,H., Aigle,M., Aljinovic,G., Andre,B., Bactet,M.C., Barthe,C., Baur,A., Becam,A.M., Bileau,N., Boles,E., Brandt,T., Brendel,M., Bruckner,M., Busserreau,F., Christiansen,C., Contreiras,R., Crouzet,M., Czapluch,C., Demolis,N., Delvaux,T., Dolignon,F., Domdey,H., Duesterhus,S., Dubois,E., Dujon,B., El Bakoury,M., Enliat,K.D., Feuermann,M., Fiers,W., Fobo,G.M., Fritz,C., Gassenhuber,H., Glansdorff,N., Goifeau,A., Griwall,L.A., de Haan,M., Hein,C., Herbert,C.J., Hollenberg,C.P., Holmstrom,K., Jacq,C., Jacques,M., Janniaux,J.C., Jomiaux,J.L., Kallioe,T., Klesau,P., Kirchath,L., Koetter,P., Korol,S., Lieb,S., Logghe,M., Lohan,A.J.E., Louis,E.U., Li,Z.Y., Mat,M.T., Mallet,L., Manhaupt,G., Messenguy,F., Miosga,T., Molemans,F., Mueller,S., Nasr,F., Obermaier,B., Pereira,J., Pierard,A., Piravandi,E., Rad,M., Rieger,M., Rose,M., Schaaf-Gerstenshaeger,I., Scherens,B., Schwarlose,C., Skala,J., Slonimski,P.P., Smits,P.H.M., Souciet,J.L., Steensma,H.Y., Stucka,R., Urestarazu,A., van der Aart,Q.J., van Dyck,A., Vassarotti,A., Vetter,I., Vierendeels,F., Vissers,S., Wagner,G., de Weyffosse,P., Wolte,K.H., Zagulski,M., Zimmermann,F.K., Mewes,H.W. and Kleine,K.	European yeast chromosome II sequencing project. MIPS at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 16a D-82152 Martinsried, FRG; E-mail: Mewes@mps.emblnet.org
FEATURES	source	location/qualifiers
gene	1..2258	/organism="Saccharomyces cerevisiae"
CDs	/db_xref="taxon:4932"	
	/chromosome="II"	
	307..2031	
	/gene="RFT1"	
	307..2031	
	/gene="RFT1"	
	/note="ORF YBL020w"	
	/codon_start=1	
	/db_xref="PID:g536015"	
	/db_xref="SWISS-PROT:P38206"	
	/translation="MAKNSQLPSTSPQSLERSTGATFLMGOLFRTKVFINNLLIRLSPRIGTIAFLPIGCTVLPFSQDAIRLSTLRISDSNGIIDDDEFEQETIRKSVLODAVFAITPEWIGFPLSLGLTAMQIRNINAFITLPEFRMSIFILWSIIVEKLSPEFTVQVQMLNTAARSFESIAVTTCIYVIVYVQOVSFPGVATSDIDEGLIAPALGRLKHSITLLACQYWDYKRNPKPLSTRLTKIKTRNNELKGYPKRSITSEFQNDIQHKEKVFQLCFKHLLEGKRLINSCVTEQGIIVALSNGSLRLRLAPAEESRLFLARLISHNPNKRLSEVILNPRFVITYSIMILVGNASSFLIQLGISKWTSTVLDTRIVRYCFYIPLSLNGPEAFQSVANGSDQILKSYFPMARSGITLMSLLIEKTKLSIEGLISNINMTRILTYGVFLNKRHRELFTSSPFENKDFETVILAGSTICLLDMWFITGVYKNDQFVNVLFAMGLALLILVKERQTIQSPINRRAVNSNDV"	
BASE COUNT	ORIGIN	
655 a	416 c	441 g
		746 t
Query Match	Best Local Similarity	Score 90; DB 24; Length 2258;
Matches 14; Conservative	5; Mismatches 4; Indels 0; Gaps 0;	
Db 396	ACTTTTCACCAAACTGTAACGT 418	
QY 1	ACNTYTCAYCARAAVGNAAVGT 23	
RESULT 14	LOCUS	PLN
SCU15087	2551 bp	31-JAN-1996
DEFINITION	Saccharomyces cerevisiae nuclear division (RFT1) gene, complete cds.	
ACCESSION	U15087	
NID	9558439	
KEYWORDS		
SOURCE	baker's yeast.	
ORGANISM	Saccharomyces cerevisiae	

THIS PAGE BLANK (USPTO)

 NWSENF
 (TM)

Release 3.1a John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPerch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Fri Apr 16 14:25:48 1999; Maspar time 3.31 Seconds
 Tabular output not generated. 90.655 Million cell updates/sec

Title: >US-09-011-797-4
 Description: (1-8) from US09011797.pep
 Perfect Score: 53
 Sequence: 1 TLHONGNV 8

Scoring table: PAM 150
 Gap 15

Searched: 116695 segs, 37453910 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: plr58
 1:plr1 2:plr2 3:plr3 4:plr4

Statistics: Mean 19.697; Variance 20.848; scale 0.945

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	53	100.0	109	2	S70631	1.31e-02
2	53	100.0	168	2	S60337	1.31e-02
3	53	100.0	176	2	JC6152	1.31e-02
4	53	100.0	181	2	JC6151	1.31e-02
5	53	100.0	187	2	JC4502	1.31e-02
6	45	84.9	1051	2	A38373	1.54e+00
7	44	83.0	384	2	B64940	2.71e+00
8	43	81.1	251	2	F64559	4.73e+00
9	43	81.1	636	2	JC4960	4.73e+00
10	41	77.4	112	2	E71313	1.40e+01
11	41	77.4	224	2	S20463	1.40e+01
12	41	77.4	261	2	B64783	1.40e+01
13	41	77.4	508	2	S54021	1.40e+01
14	41	77.4	726	2	S22258	1.40e+01
15	41	77.4	863	2	S41984	1.40e+01
16	40	75.5	240	2	S71458	2.37e+01
17	40	75.5	240	2	S62088	2.37e+01
18	40	75.5	250	2	S70854	2.37e+01
19	40	75.5	298	2	A64922	2.37e+01
20	40	75.5	328	2	D69452	2.37e+01
21	40	75.5	639	2	A39135	2.37e+01
22	40	75.5	745	2	B70017	2.37e+01
23	39	73.6	74	2	S72807	3.98e+01

RESULT	ENTRY	DESCRIPTION	SCORE	ALIGNMENTS		
24	39	73.6	147	2	B48402	Sigma factor F inhibi
25	39	73.6	149	2	E41317	probable lipopolysacc
26	39	73.6	158	2	S34612	anik protein homolog
27	39	73.6	219	2	B47152	probable fibrinogen
28	39	73.6	232	2	S47846	rfay protein - Escher
29	39	73.6	398	2	E71539	hypothetical protein
30	39	73.6	412	2	G64736	YadC protein - Escher
31	39	73.6	439	2	F69281	NADH oxidase (nox-1)
32	39	73.6	462	2	D32057	nitrogenase (EC 1.18.
33	39	73.6	720	2	S61143	KRE6 protein - yeast
34	39	73.6	723	2	A36481	myosin homolog, brain
35	39	73.6	885	2	I38968	100 kDa coactivator -
36	39	73.6	892	2	S63378	hypothetical protein
37	39	73.6	971	1	JQ1634	outer capsid protein
38	38	71.7	238	2	D69021	N5-methyl-tetrahydrom
39	38	71.7	325	2	F70440	conserved hypothetical
40	38	71.7	533	2	JC1103	lens fiber cell beade
41	38	71.7	601	2	E71435	probable triacylglyce
42	38	71.7	634	2	C64454	hypothetical protein
43	38	71.7	893	2	S46442	nitrate reductase (NA
44	38	71.7	941	2	A55195	chordin precursor - A
45	38	71.7	1319	2	C43735	besc protein - Acetob

RESULT 1
 ENTRY S70631 #type fragment
 TITLE orphanin FQ/nociceptin - mouse (fragment)
 ORGNISM #formal_name Mus musculus #common_name house mouse
 DATE 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 10-Sep-1997

ACCESSIONS S70631
 REFERENCE S70631
 #authors Pan, Y.X.; Xu, J.; Pasternak, G.W.
 #journal Biochem. J. (1996) 315:11-13
 #title Cloning and expression of a cDNA encoding a mouse brain orphanin FQ/nociceptin precursor.
 #accession S70631
 #status preliminary
 #molecule_type mRNA
 #residues 1-109 #label PAN
 #cross-references EMBL:U44027; NID:91335869; PID:91335870
 #length 109 #checksum 9509

SUMMARY
 Query Match 100.0%; Score 53; DB 2; Length 109;
 Best Local Similarity 100.0%; Pred. No. 1.31e-02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB	Score	Query Match	Length	DB ID	Description	Pred. No.
Db	102	TLHONGNV	109			
Qy	1	TLHONGNV	8			

RESULT 2
 ENTRY S60337 #type fragment
 TITLE oploild receptor-like ORL(1) receptor antagonist - rat (fragment)
 ORGNISM #formal_name Rattus norvegicus #common_name Norway rat
 DATE 06-Sep-1996 #sequence_revision 27-Feb-1997 #text_change 18-Mar-1997

ACCESSIONS S60337
 REFERENCE S60337
 #authors Meunier, J.C.; Mollereau, C.; Toll, L.; Snaudreu, C.; Ferrera, C.; Alvinerie, P.; Butour, J.L.; Guillemot, J.C.; Parmentier, M.; Costentin, J.
 #journal Nature (1995) 377:532-535
 #title Isolation and structure of the endogenous agonist of oploild receptor-like ORL(1) receptor.
 #accession S60337
 #status preliminary; not compared with conceptual translation
 #molecule_type mRNA


```

#residues 1-168 ##label MEU
SUMMARY #length 168 #checksum 3964

Query Match 100.0%; Score 53; DB 2; Length 168;
Best Local Similarity 100.0%; Pred. No. 1.31e-02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 161 TLHONGNV 168
| | | | | | | |
| 1 TLHONGNV 8

RESULT 3
ENTRY JC6152 #type complete
TITLE orphanin FQ precursor - human
ALTERNATE_NAMES nociceptin precursor
ORGANISM #formal_name Homo sapiens #common_name man
DATE 11-Apr-1997 #sequence_revision 09-May-1997 #text_change
10-Sep-1997

ACCESSIONS JC6152
REFERENCE JC6151
AUTHORS Nothacker, H.P.; Reinscheid, R.K.; Mansour, A.; Henningsen,
R.A.; Ardati, A.; Monsma Jr., F.J.; Watson, S.J.; Civelli,
O.
#journal Proc. Natl. Acad. Sci. U.S.A. (1996) 93:8677-8682
#title Primary structure and tissue distribution of the orphanin FQ
#accession JC6152
#molecule_type mRNA
#residues 1-176 ##label NOT
#cross-references GB:U48263; NID:g1185009; PID:g1185010
COMMENT This protein is a neuropeptide that is an endogenous ligand to a
G-protein-coupled receptor sequentially related to the opioid
receptors. It functions in development.
KEYWORDS opioid peptide
SUMMARY #length 176 #molecular_weight 20295 #checksum 498

Query Match 100.0%; Score 53; DB 2; Length 176;
Best Local Similarity 100.0%; Pred. No. 1.31e-02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 169 TLHONGNV 176
| | | | | | | |
| 1 TLHONGNV 8

RESULT 4
ENTRY JC6151 #type complete
TITLE orphanin FQ precursor - rat
ALTERNATE_NAMES nociceptin precursor
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 11-Apr-1997 #sequence_revision 09-May-1997 #text_change
10-Sep-1997

ACCESSIONS JC6151
REFERENCE JC6151
AUTHORS Nothacker, H.P.; Reinscheid, R.K.; Mansour, A.; Henningsen,
O.; Ardati, A.; Monsma Jr., F.J.; Watson, S.J.; Civelli,
R.A.; Ardati, A.; Monsma Jr., F.J.; Watson, S.J.; Civelli,
O.
#journal Proc. Natl. Acad. Sci. U.S.A. (1996) 93:8677-8682
#title Primary structure and tissue distribution of the orphanin FQ
#accession JC6151
#molecule_type mRNA
#residues 1-181 ##label NOT
#cross-references GB:U48262; NID:g1185011; PID:g1185012
COMMENT This protein is a neuropeptide that is an endogenous ligand to a
G-protein-coupled receptor sequentially related to the opioid
receptors. It is involved in development.
KEYWORDS opioid peptide
SUMMARY #length 181 #molecular_weight 20478 #checksum 3303

Query Match 100.0%; Score 53; DB 2; Length 181;
Best Local Similarity 100.0%; Pred. No. 1.31e-02;

```

```

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 174 TLHONGNV 181
| | | | | | | |
| 1 TLHONGNV 8

RESULT 5
ENTRY JC4502 #type complete
TITLE opioid-receptor like G protein coupled receptor
ALTERNATE_NAMES nociceptin/orphanin FQ protein precursor - mouse
ORGANISM N23K protein; nociceptin precursor homolog
#formal_name Mus musculus #common_name house mouse
DATE 07-Feb-1996 #sequence_revision 19-Apr-1996 #text_change
10-Sep-1997

ACCESSIONS JC4502
REFERENCE JC4502
AUTHORS Saito, Y.; Maruyama, K.; Saito, T.C.; Kawashima, S.
#journal Biochem. Biophys. Res. Commun. (1995) 217:539-545
#title N23K, a gene transiently up-regulated during neural
differentiation, encodes a precursor protein for a newly
identified neuropeptide nociceptin.
#accession JC4502
#molecule_type mRNA
#residues 1-187 ##label SAT
#cross-references DBJ:D50056; NID:g1857023; PID:d1009395; PID:g1217906
REFERENCE JC4502
AUTHORS Houtani, T.; Nishi, M.; Takeshima, H.; Nukada, T.; Sugimoto,
T.
#journal Biochem. Biophys. Res. Commun. (1996) 219:714-719
#title Structure and regional distribution of nociceptin/orphanin FQ
precursor.
#accession JC4502
#molecule_type mRNA
#residues 1-187 ##label HOI
#cross-references DBJ:D52866; NID:g1311472; PID:d1012281; PID:g1311473
#experimental_source brain
#note The authors translated the codon CTC for residue 72 as
lys and GAC for residue 76 as Glu
COMMENT This protein belongs to the opiate precursor family, and functions
as a neuropeptide precursor and plays a role in neuronal
differentiation and development.
GENETICS N23K
#gene brain; differentiation; G protein-coupled receptor;
#keywords neuropeptide
FEATURE 1-18
19-187
109-126 #domain signal sequence #status predicted #label SIG\
#label MAT\
#region D-A-E-P-G-A motif\
141-157 #product neuropeptide #status predicted #label NRP\
160-176 #product heptadecapeptide #status predicted #label HRP
SUMMARY #length 187 #molecular_weight 20864 #checksum 8738

Query Match 100.0%; Score 53; DB 2; Length 187;
Best Local Similarity 100.0%; Pred. No. 1.31e-02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 180 TLHONGNV 187
| | | | | | | |
| 1 TLHONGNV 8

RESULT 6
ENTRY A38373 #type complete
TITLE ubiquitin--protein ligase (EC 6.3.2.19) E1 - wheat
ALTERNATE_NAMES ubiquitin-activating enzyme E1, UBA1
ORGANISM #formal_name Triticum sp. #common_name wheat
DATE 24-Jul-1991 #sequence_revision 28-Jun-1991 #text_change
24-Jul-1998
ACCESSIONS A38373; A42873

```



```

REFERENCE      A38373
#authors      Hatfield, P.M.; Callis, J.; Vierstra, R.D.
#journal      J. Biol. Chem. (1990) 265:15813-15817
#title        Cloning of ubiquitin activating enzyme from wheat and
              expression of a functional protein in Escherichia coli.
#cross-references NID:90368797
#accession    A38373
#status       preliminary
#molecule_type mRNA
#residues     1-1051 #label HMT
#cross-references GB:M55604; GB:M33631; NID:g170779; PID:g170780
REFERENCE      A42873
#authors      Hatfield, P.M.; Vierstra, R.D.
#journal      J. Biol. Chem. (1992) 267:14799-14803
#title        Multiple forms of ubiquitin-activating enzyme E1 from wheat.
              Identification of an essential cysteine by in vitro
              mutagenesis.
#cross-references NID:92340519
#accession    A42873
#status       preliminary; not compared with conceptual translation
#molecule_type mRNA
#residues     1-1051 #label HMT
#note         sequence extracted from NCBI backbone (NCBIP:109415)
CLASSIFICATION #superfamily ubiquitin-activating enzyme E1
KEYWORDS       ligase
SUMMARY        #length 1051 #molecular-weight 117007 #checksum 9170

Query Match      84.9%; Score 45; DB 2; Length 1051;
Best Local Similarity 75.0%; Pred. No. 1.54e+00;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 93 TLHDDGNY 100
QY 1 TLHQNGNV 8

RESULT 7
ENTRY   B64940 #type complete
TITLE   hypothetical protein b1794 - Escherichia coli (strain K-12)
ORGANISM #formal_name Escherichia coli
DATE     12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change
14-Nov-1997
ACCESSIONS B64940
REFERENCE   A64720
#authors   Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
            Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
            Rode, C.K.; Mayhew, G.F.; Collado-Vides, J.; Glasner, J.D.;
            Kirpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
            Y.
#journal   Science (1997) 277:1453-1462
#title     The complete genome sequence of Escherichia coli K-12.
#cross-references NID:97426517
#accession B64940
#status     preliminary; nucleic acid sequence not shown;
            translation not shown
#molecule_type DNA
#residues   1-384 #label BLAT
#cross-references GB:AE000274; GB:U00096; NID:g1788089; PID:g1788095;
            UMGF:b1794
SUMMARY     #experimental_source strain K-12, substrain MG1655
            #length 384 #molecular-weight 43628 #checksum 3691

Query Match      83.0%; Score 44; DB 2; Length 384;
Best Local Similarity 75.0%; Pred. No. 2.71e+00;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 284 TLQONGEV 291
QY 1 TLHQNGNV 8

RESULT 8
ENTRY   F64559 #type complete

```

```

TITLE          conserved hypothetical protein HP0318 - Helicobacter pylori
              (strain 26695)
ORGANISM       #formal_name Helicobacter pylori
DATE           09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change
05-Jun-1998
ACCESSIONS     F64559
REFERENCE      A64520
#authors       Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.;
              Sutton, G.G.; Fleischmann, R.D.; Ketchum, K.A.; Klein,
              H.P.; Gill, S.; Dougherty, B.A.; Nelson, K.; Quackenbush,
              J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus, B.;
              Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.;
              McKenney, K.; Fitzgerald, L.M.; Lee, N.; Adams, M.D.;
              Hickey, E.K.; Berg, D.E.; Gocayne, J.D.; Utterback, J.R.;
              Peterson, C.; Bowman, C.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.;
              Fujii, C.; Bowman, C.; Wathey, L.; Wallin, E.; Hayes,
              W.S.; Borodovsky, M.; Karp, P.D.; Smith, H.O.; Fraser,
              C.M.; Venter, J.C.
#journal       Nature (1997) 388:539-547
#title         The complete genome sequence of the gastric pathogen
              Helicobacter pylori.
#cross-references NID:97394467
#accession    F64559
#status       preliminary; nucleic acid sequence not shown;
              translation not shown
#molecule_type DNA
#residues     1-251 #label TOM
#cross-references GB:AE000550; GB:AE000511; NID:g2313417; PID:g2313418;
            TIGR:HP0318
CLASSIFICATION #superfamily conserved hypothetical protein HP0318
SUMMARY        #length 251 #molecular-weight 28507 #checksum 8065

Query Match      81.1%; Score 43; DB 2; Length 251;
Best Local Similarity 75.0%; Pred. No. 4.73e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 105 TLHPNGHV 112
QY 1 TLHQNGNV 8

RESULT 9
ENTRY   JC4960 #type complete
TITLE   DNA topoisomerase (ATP-hydrolyzing) (EC 5.99.1.3) B -
            Thermotoga maritima
ALTERNATE_NAMES DNA gyrase B like; type II DNA topoisomerase
ORGANISM       #formal_name Thermotoga maritima
DATE           31-Dec-1996 #sequence_revision 31-Dec-1996 #text_change
08-Sep-1997
ACCESSIONS     JC4960
REFERENCE      JC4960
#authors       Guipaud, O.; Labedan, B.; Forterre, P.
#journal       Gene (1996) 174:121-128
#title         A gybB-like gene from the hyperthermophilic bacterium
            Thermotoga maritima.
#accession    JC4960
#molecule_type DNA
#residues     1-636 #label GVI
#cross-references GB:U49692; NID:g1622790; PID:g1622792
COMMENT        This enzyme is an adenosine triphosphate dependent enzyme with DNA
            relaxation and decatenation activities.
GENETICS       top2B
#gene          #superfamily DNA topoisomerase (ATP-hydrolyzing) chain B
CLASSIFICATION isomerase
KEYWORDS        #length 636 #molecular-weight 72459 #checksum 3136
SUMMARY

Query Match      81.1%; Score 43; DB 2; Length 636;
Best Local Similarity 62.5%; Pred. No. 4.73e+00;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 60 TLHDDGSV 67
QY 1 TLHQNGNV 8

```



```

QY      1 TLHQNGNV 8

RESULT  10
ENTRY   E71313
TITLE   Probable V-type ATPase, subunit F - syphilis spirochete
ORGANISM #formal_name Treponema pallidum subsp. pallidum #common_name
          syphilis spirochete
DATE    24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change
          07-Aug-1998

ACCESSIONS
REFERENCE E71313
          A71250
          Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.;
          Sutton, G.G.; Dodson, R.; Gwin, M.; Hickey, E.K.; Clayton,
          R.; Ketchum, K.A.; Sodergren, E.; Hardham, J.M.; McLeod,
          M.P.; Salzberg, S.; Petersen, J.; Khalak, H.; Richardson,
          D.; Howell, J.K.; Chidambaram, M.; Uterback, T.; McDonald,
          L.; Artlach, P.; Bowman, C.; Cotton, M.D.; Fujii, C.;
          Garland, S.; Hatch, B.; Horst, K.; Roberts, K.; Watthey,
          L.; Weidman, J.; Smith, H.O.; Venter, J.C.
          Science (1998) 281:375-388

#journal
#title   Complete genome sequence of Treponema pallidum, the syphilis
          spirochete.
#accession E71313
#status   preliminary; nucleic acid sequence not shown;
          translation not shown
#molecule_type DNA
#residues 1-112 #label COL
#cross-references GB:AE001228; GB:AE000520; NID:g3322816; PID:g3322823
#experimental_source strain Nichols

GENETICS
SUMMARY  TP0531 #length 112 #molecular-weight 12494 #checksum 6378
          77.4%; Score 41; DB 2; Length 112;
          Best Local Similarity 62.5%; Pred. No. 1.40e+01;
          Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db      89 ALHOGQNV 96
          :||: |||
          QY      1 TLHQNGNV 8

RESULT  11
ENTRY   S20463
TITLE   siderophore biosynthesis regulatory protein sfp - Bacillus
          subtilis (strain OXB105)
ALTERNATE_NAMES sfp
ORGANISM #formal_name Bacillus subtilis
DATE    13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change
          18-Sep-1998

ACCESSIONS
REFERENCE S20463
          Nakano, M.M.; Corbelli, N.; Besson, J.; Zuber, P.
          Mol. Gen. Genet. (1992) 232:313-321
          Isolation and characterization of sfp: a gene that functions
          in the production of the lipopeptide biosurfactant,
          surfactin, in Bacillus subtilis.
          #cross-references MUID:92212296
          #accession S20463
          #molecule_type DNA
          #residues 1-224 #label NAK
          #cross-references EMBL:X63158; NID:g40138; PID:g40139
          #experimental_source strain OXB105
COMMENT   This protein (sfp variant) is found in surfactin producing strains.
          Strains that do not produce surfactin have a sfp(0) variant with
          a carboxyl-terminal sequence that differs from this sequence due
          to a single nucleotide deletion and frameshift. The sequence of a
          sfp(0) variant can be found in FIR:A36931.

GENETICS
#gene     sfp; sfp(0)
CLASSIFICATION #superfamily siderophore biosynthesis regulatory protein sfp
KEYWORDS   antibiotic biosynthesis

SUMMARY  #length 224 #molecular-weight 26135 #checksum 6728
          Query Match
          Best Local Similarity 71.4%; Score 41; DB 2; Length 224;
          Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db      172 LHQDQGV 178
          |||||
          QY      2 LHQNGNV 8

RESULT  12
ENTRY   B64783
TITLE   hypothetical protein b0515 - Escherichia coli
ORGANISM #formal_name Escherichia coli
DATE    12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change
          17-Jul-1998

ACCESSIONS
REFERENCE B64783
          A64720
          Blatner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
          Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
          Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
          Kirkpatrick, H.A.; Goeden, M.A.; Kose, D.V.; Mau, B.; Shao,
          Y.
          Science (1997) 277:1453-1462
          #title   The complete genome sequence of Escherichia coli K-12.
          #cross-references MUID:97426617
          #accession B64783
          #status   nucleic acid sequence not shown; translation not shown
          #molecule_type DNA
          #residues 1-261 #label BLAT
          #cross-references GB:AE000157; GB:U00096; NID:g1786716; PID:g1786725;
          UMG:P:00515

SUMMARY  #experimental_source strain K-12, substrain MG1655
          #length 261 #molecular-weight 28730 #checksum 3167

Query Match
Best Local Similarity 100.0%; Score 41; DB 2; Length 261;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      65 TLHQNG 70
          |||||
          QY      1 TLHQNG 6

RESULT  13
ENTRY   S54021
TITLE   FMS1 protein - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES protein YW9711.09; protein YW920W
ORGANISM #formal_name Saccharomyces cerevisiae
DATE    08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change
          24-Sep-1998

ACCESSIONS
REFERENCE S54021; S67323; S72321
          Lye, G.; Churcher, C.M.
          Submitted to the EMBL Data Library, May 1995
          #accession S54021
          #molecule_type DNA
          #residues 1-508 #label LYE
          #cross-references EMBL:Z49211; NID:g798922; PID:g798930; MIPS:YWR020W
          #experimental_source strain AB972
          S67323
          Pousset, D.; Marcilieu, M.; Karst, F.
          Submitted to the EMBL Data Library, September 1994
          #description Characterisation of FMS1, a multicopy suppressor of
          farnoproporph resistance in yeast.
          #accession S67323
          #molecule_type DNA
          #residues 1-508 #label POU
          #cross-references EMBL:X81848; NID:g1143555; PID:g113924; PID:g1143556
          S72321
          Joets, J.; Pousset, D.; Marcilieu, C.; Karst, F.
          Curr. Genet. (1996) 30:115-120

```



```

#title      Characterization of the Saccharomyces cerevisiae FMS1 gene
            related to Candida albicans corticosteroid-binding protein
            1.
#accession  S72321
#status     nucleic acid sequence not shown
#molecule_type DNA
#residues   1-508 ##label JOE
#cross-references EMBL:X81846; NID:g1143555; PID:e119324; PID:g1143556
GENETICS
#gene       SGD:FMS1
#cross-references SGD:S0004622; MIPS:YMR020w
FUNCTION
#map_position 13R
#description  multicopy suppressor of fenpropimorph resistance
KEYWORDS     transmembrane protein
FEATURE
11-27        #domain transmembrane #status predicted #label TMM
SUMMARY      #length 508 #molecular-weight 57805 #checksum 4542

Query Match      77.48; Score 41; DB 2; Length 508;
Best Local Similarity 100.0%; Pval. NO. 1.40e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      26 TLHONG 31
        |||||
Oy      1 TLHONG 6

RESULT 14
ENTRY   S22258      #type complete
TITLE   Probable protein kinase KIN82 (EC 2.7.1.-) - yeast
         (Saccharomyces cerevisiae)
ALTERNATE_NAMES
ORGANISM protein_YCR091w
DATE     #format_name Saccharomyces cerevisiae
         04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change
         10-Jul-1998
ACCESSIONS
REFERENCE S22258; S19507; S19769
#authors  Wilson, C.; Bergantino, E.; Lanfranchi, G.; Valle, G.;
         Carignani, G.; Frontali, L.
#journal  Yeast (1992) 8:71-77
#title    A putative serine/threonine protein kinase gene on chromosome
         III of Saccharomyces cerevisiae.
#cross-references MIMD:92254506
#accession S22258
#molecule_type DNA
#residues 1-726 ##label MIL
#cross-references EMBL:X5720; NID:g1907116; PID:e264713; PID:g1907229
REFERENCE S19504
#authors  Frontali, L.; Grisanti, P.
#submission Submitted to the Protein Sequence Database, March 1992
#accession S19507
#molecule_type DNA
#residues 1-580 ##label PRO
#cross-references EMBL:X59720; MIPS:YCR091w
REFERENCE S19508
#authors  Bergantino, E.; Carignani, G.; Lanfranchi, G.; Valle, G.
#submission Submitted to the Protein Sequence Database, March 1992
#accession S19769
#molecule_type DNA
#residues 292-726 ##label BER
#cross-references EMBL:X59720; MIPS:YCR091w
GENETICS
#gene       SGD:KIN82
#cross-references SGD:S0000667; MIPS:YCR091w
#map_position 3R
CLASSIFICATION #superfamily unassigned Ser/Thr or Tyr-specific protein
               kinases; protein kinase homology
KEYWORDS     ATP; phosphotransferase; serine/threonine-specific protein
               kinase
FEATURE
322-662     #domain protein kinase homology #label KIN\
350-358     #region protein kinase ATP-binding motif\

```

```

353                               #active_site lys #status predicted
SUMMARY                          #length 726 #molecular-weight 82122 #checksum 5589

Query Match                      77.4%; Score 41; DB 2; Length 726;
Best Local Similarity            71.4%; Pred. No. 1.40e+01;
Matches                          5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db      457 LHONGNV 463
      |||:|:|
QY      2 LHONGNV 8

RESULT  15
ENTRY   S41984 #type complete
TITLE   SIPI protein - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES  protein YDR422c
ORGANISM #formal_name Saccharomyces cerevisiae
DATE    31-Mar-1992 #sequence_revision 14-Sep-1994 #text_change
        06-Feb-1998
ACCESSIONS  S41984; S69705
REFERENCE   S41984
#authors   Yang, X.; Hubbard, E.J.A.; Carlson, M.
#journal   Science (1992) 257:680-682
#title     A protein kinase substrate identified by the two-hybrid
          system.
#accession S41984
#status    nucleic acid sequence not shown
#molecule_type DNA
#residues  1-863 #label YAN
#cross-references EMBL:M90531; NID:q172603; PID:q172604
REFERENCE  S69553
#authors   Dietrich, F.S.
#submission submitted to the EMBL Data Library, August 1995
#description The sequence of S. cerevisiae lambda 3641 and cosmids 9461,
          9831, and 9410.
#accession S69705
#molecule_type DNA
#residues  1-863 #label DIE
#cross-references EMBL:U33007; NID:g927685; PID:g927721; MIPS:YDR422c
GENETICS
#gene       SGD:SIPI
#map_position 4R
#cross-references SGD:S0002830; MIPS:YDR422c
SUMMARY   #map_position 4R #length 863 #molecular-weight 96258 #checksum 2426

Query Match                      77.4%; Score 41; DB 2; Length 863;
Best Local Similarity            62.5%; Pred. No. 1.40e+01;
Matches                          5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db      754 TLQOQGN 761
      |||:|:|
QY      1 TLQOQGNV 8

Search completed: Fri Apr 16 14:26:05 1999
Job time : 17 secs.

```


THIS PAGE BLANK (USPTO)


```

##residues 1-168 ##label MEU
SUMMARY #length 168 #checksum 3964

Query Match 100.0%; Score 131; DB 2; Length 168;
Best Local Similarity 100.0%; Pred. No. 9.72e-13;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 141 FSEFMROYLVLSMOSSQ 157
QY 1 FSEFMROYLVLSMOSSQ 17

RESULT 3
ENTRY JC6152 #type complete
TITLE orphanin FQ precursor - human
ALTERNATE_NAMES nociceptin precursor
ORGANISM #formal_name Homo sapiens #common_name man
DATE 11-Apr-1997 #sequence_revision 09-May-1997 #text_change
10-Sep-1997

ACCESSIONS JC6152
REFERENCE JC6151
#authors Nothacker, H.P.; Reinscheid, R.K.; Mansour, A.; Henningsen,
R.A.; Ardati, A.; Monsma Jr., F.J.; Watson, S.J.; Civelli,
O.
#journal Proc. Natl. Acad. Sci. U.S.A. (1996) 93:8677-8682
#title Primary structure and tissue distribution of the orphanin FQ
#accession precursor.
#molecule_type mRNA
#residues 1-176 ##label NOT
#cross-references GB:U48263; NID:q1185009; PID:q1185010
COMMENT This protein is a neuropeptide that is an endogenous ligand to a
G-protein-coupled receptor sequentially related to the opiod
receptors. It functions in development.
opiod peptide
KEYWORDS #length 176 #molecular-weight 20295 #checksum 498
SUMMARY

Query Match 100.0%; Score 131; DB 2; Length 176;
Best Local Similarity 100.0%; Pred. No. 9.72e-13;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 149 FSEFMROYLVLSMOSSQ 165
QY 1 FSEFMROYLVLSMOSSQ 17

RESULT 4
ENTRY JC6151 #type complete
TITLE orphanin FQ precursor - rat
ALTERNATE_NAMES nociceptin precursor
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 11-Apr-1997 #sequence_revision 09-May-1997 #text_change
10-Sep-1997

ACCESSIONS JC6151
REFERENCE JC6151
#authors Nothacker, H.P.; Reinscheid, R.K.; Mansour, A.; Henningsen,
R.A.; Ardati, A.; Monsma Jr., F.J.; Watson, S.J.; Civelli,
O.
#journal Proc. Natl. Acad. Sci. U.S.A. (1996) 93:8677-8682
#title Primary structure and tissue distribution of the orphanin FQ
#accession precursor.
#molecule_type mRNA
#residues 1-181 ##label NOT
#cross-references GB:U48262; NID:q1185011; PID:q1185012
COMMENT This protein is a neuropeptide that is an endogenous ligand to a
G-protein-coupled receptor sequentially related to the opiod
receptors. It is involved in development.
opiod peptide
KEYWORDS #length 181 #molecular-weight 20478 #checksum 3303
SUMMARY

Query Match 100.0%; Score 131; DB 2; Length 181;
Best Local Similarity 100.0%; Pred. No. 9.72e-13;

```

```

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 154 FSEFMROYLVLSMOSSQ 170
QY 1 FSEFMROYLVLSMOSSQ 17

RESULT 5
ENTRY JC4502 #type complete
TITLE opiod-receptor like G protein coupled receptor
ALTERNATE_NAMES nociceptin/orphanin FQ protein precursor - mouse
ORGANISM N3K protein; nociceptin precursor homolog
#formal_name Mus musculus #common_name house mouse
DATE 07-Feb-1996 #sequence_revision 19-Apr-1996 #text_change
10-Sep-1997

ACCESSIONS JC4502; JC4652
REFERENCE JC4502
#authors Saito, Y.; Maruyama, K.; Saito, T.C.; Kawashima, S.
#journal Biochem. Biophys. Res. Commun. (1995) 217:539-545
#title N3K, a gene transiently up-regulated during neural
differentiation, encodes a precursor protein for a newly
identified neuropeptide nociceptin.
#accession JC4502
#molecule_type mRNA
#residues 1-187 ##label SAI
#cross-references DDBJ:D50056; NID:q1857023; PID:d1009395; PID:g1217906
REFERENCE JC4652
#authors Houtant, T.; Nishi, M.; Takeshima, H.; Nukada, T.; Sugimoto,
T.
#journal Biochem. Biophys. Res. Commun. (1996) 219:714-719
#title Structure and regional distribution of nociceptin/orphanin FQ
precursor.
#accession JC4652
#molecule_type mRNA
#residues 1-187 ##label HOU
#cross-references DDBJ:D82866; NID:q1311472; PID:d1012281; PID:g1311473
#experimental_source brain
#note The authors translated the codon CRC for residue 72 as
Lys and GAC for residue 76 as Glu
COMMENT This protein belongs to the opiate precursor family, and functions
as a neuropeptide precursor and plays a role in neuronal
differentiation and development.

GENETICS N3K
#gene brain; differentiation; G protein-coupled receptor;
#keywords neuropeptide
FEATURE 1-18
19-187
109-126 #domain signal sequence #status predicted #label SIG\
141-157 #product neuropeptide precursor, N3K #status predicted
160-176 #label MAP\
#region D-A-E-P-G-A motif\
#product neuropeptide #status predicted #label NRP\
#product heptadecapeptide #status predicted #label HPP
SUMMARY #length 187 #molecular-weight 20884 #checksum 8758

Query Match 100.0%; Score 131; DB 2; Length 187;
Best Local Similarity 100.0%; Pred. No. 9.72e-13;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 160 FSEFMROYLVLSMOSSQ 176
QY 1 FSEFMROYLVLSMOSSQ 17

RESULT 6
ENTRY G70523 #type complete
TITLE hypothetical protein RV0302 - Mycobacterium tuberculosis
ORGANISM (strain H37RV)
#formal_name Mycobacterium tuberculosis
DATE 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
17-Jul-1998
ACCESSIONS G70523

```


REFERENCE
#anchors A70500
Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekle, E.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Hornsby, L.; Jags, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Squares, R.; Sultson, J.E.; Taylor, K.; Whitehead, S.; Barrett, B.G.
#journal Nature (1998) 393:537-544
#title Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.
#cross-references MUID:96295987
#accession G70523
#status preliminary: nucleic acid sequence not shown;
translation not shown
##molecule-type DNA
##residues 1-210 #label COL
##cross-references GB:296800; GB:AL123456; NID:g261800; PID:e321660; PID:g2193923
##experimental_source strain H37Rv
GENETICS
#gene Rv0302
SUMMARY #length 210 #molecular-weight 23396 #checksum 5675
Query Match 51.9%; Score 68; DB 2; Length 210;
Best Local Similarity 43.8%; Pred. No. 1.19e+00;
Matches 7; Conservative 6; Mismatches 3; Indels 0; Gaps 0;
DB 106 PDFRLFLYLSMERSQ 121
QY 2 SEFMROYLVLSMOSSQ 17
RESULT 7
ENTRY D70139 #type complete
TITLE conserved hypothetical integral membrane protein BB0317 - Lyme disease spirochete
ORGANISM #formal_name Borrelia burgdorferi #common_name Lyme disease spirochete
DATE 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 05-Jun-1998
ACCESSIONS D70139
REFERENCE A70100
#anchors Fraser, C.M.; Castens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White, O.; Ketchum, K.A.; Dodson, R.; Hickey, E.K.; Gwinn, M.; Dougherty, B.; Tomb, J.F.; Fleischmann, R.D.; Richardson, D.; Peterson, J.; Kierlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vagt, R.V.; Palmer, N.; Adams, M.D.; Gocayne, J.; Weidman, J.; Uterback, T.; Matthey, L.; McDonald, L.; Artach, P.; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.; Smith, H.O.; Venter, J.C.
#journal Nature (1997) 390:580-586
#title Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
#cross-references MUID:98065943
#accession D70139
#status preliminary: nucleic acid sequence not shown;
translation not shown
##molecule-type DNA
##residues 1-310 #label KLE
##cross-references GB:AE001138; GB:AE000783; NID:g2688210; PID:g2688213; TIGR:BB0317
SUMMARY #experimental_source strain B31
#length 310 #molecular-weight 35680 #checksum 9426
Query Match 51.9%; Score 68; DB 2; Length 310;
Best Local Similarity 50.0%; Pred. No. 1.19e+00;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

DB 228 FVFFRPYVIGLMSG 243
QY 1 FSEFMROYLVLSMOSS 16
RESULT 8
ENTRY S42370 #type complete
TITLE citrate (sl)-synthase (EC 4.1.3.7), precursor - Caenorhabditis elegans
ORGANISM #formal_name Caenorhabditis elegans
DATE 07-Oct-1994 #sequence_revision 10-Nov-1995 #text_change 18-Sep-1998
ACCESSIONS S42370
REFERENCE S42368
#anchors Smith, A.
#submission submitted to the EMBL Data Library, March 1994
#accession S42370
#status preliminary
##molecule-type DNA
##residues 1-468 #label SMT
##cross-references EMBL:230423; NID:g458479; PID:g458482
GENETICS
#introns 69/3; 202/3; 309/3
CLASSIFICATION #superfamily citrate (sl)-synthase
KEYWORDS carbon-carbon lyase; mitochondrion; oxo-acid-lyase
SUMMARY #length 468 #molecular-weight 51540 #checksum 2659
Query Match 50.4%; Score 66; DB 2; Length 468;
Best Local Similarity 63.6%; Pred. No. 2.50e+00;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
DB 253 FAEMLRLYLV 263
QY 1 FSEFMROYLV 11
RESULT 9
ENTRY F64479 #type complete
TITLE the mononuclease precursor homolog - Methanococcus jannaschii
ORGANISM #formal_name Methanococcus jannaschii
DATE 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 10-Oct-1997
ACCESSIONS F64479
REFERENCE A64300
#anchors Bull, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, J.A.; Fitzgerald, L.M.; Clayton, R.A.; Gocayne, J.D.; Kierlavage, A.R.; Dougherty, B.A.; Tomb, J.F.; Adams, M.D.; Reisch, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Scott, J.L.; Geoghegan, N.S.M.; Weidman, J.F.; Fuhrmann, J.D.; Nguyen, D.; Uterback, T.R.; Kelley, J.M.; Peterson, K.M.; Hurst, M.A.; Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.
#journal Science (1996) 273:1058-1073
#title Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii.
#cross-references MUID:96337999
#accession F64479
#status preliminary: nucleic acid sequence not shown;
translation not shown
##molecule-type DNA
##residues 1-207 #label BUL
##cross-references GB:U67584; GB:L77117; NID:g1592077; PID:g1592087; TIGR:MJ1439; PID:g1511458
GENETICS
#map_position FOR1406535-1407158
SUMMARY #length 207 #molecular-weight 24547 #checksum 5813
Query Match 48.9%; Score 64; DB 2; Length 207;
Best Local Similarity 60.0%; Pred. No. 5.18e+00;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 6 IMRKLISM 15
:|||||
QY 4 FMRQYVLISM 13

RESULT 10
ENTRY 671131 #type complete
TITLE Probable ATP-binding protein - Pyrococcus horikoshii
ORGANISM #formal_name Pyrococcus horikoshii
DATE 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 14-Aug-1998
671131
A71000
Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.; Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.; Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Kobb, F.T.; Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.
DNA Res. (1998) 5:55-76
Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3.
#accession G71131
#status preliminary; nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-312 #label KAM
#cross-references GB:AP000003; NID:g3236130; PID:d1030856; PID:g3257230
#experimental_source strain OT3
#note this accession replaces an interim accession for a sequence replaced by GenBank

GENETICS
#gene PH0820
SUMMARY #length 312 #molecular-weight 35239 #checksum 6859

Query Match 48.9%; Score 64; DB 2; Length 312;
Best Local Similarity 69.2%; Pred. No. 5.18e+00;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 141 FSKGMROLYVLAM 153
:|||||
QY 1 FSEFMROYVLISM 13

RESULT 11
ENTRY D71472 #type complete
TITLE Probable thioredoxin disulfide isomerase - Chlamydia trachomatis (serotype D, strain UW3/Cx)
ORGANISM #formal_name Chlamydia trachomatis
DATE 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 13-Sep-1998
D71472
A71460
Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, W.P.; Olinger, L.; Yatsunov, R.L.; Zhao, Q.; Keonin, E.V.; Davis, R.W.
submitted to GenBank, May 1998
#description Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis.
#accession D71472
#status preliminary
#molecule_type DNA
#residues 1-164 #label ARN
#cross-references GB:AE001350; GB:AE001273; NID:g3329238; PID:g3329244
#experimental_source serotype D, strain UW-3/Cx

GENETICS
#gene CT780
SUMMARY #length 164 #molecular-weight 18552 #checksum 4132

Query Match 47.3%; Score 62; DB 2; Length 164;
Best Local Similarity 63.6%; Pred. No. 1.06e+01;

Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 86 FSEFAKQYICM 96
:|||||
QY 1 FSEFMROYVL 11

RESULT 12
ENTRY S66176 #type complete
TITLE ACC oxidase (clone AC03) oxidase - muskmelon
ORGANISM #formal_name Cucumis melo- #common_name muskmelon
DATE 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 13-Sep-1998
S66176
S66174
Lasserre, E.; Bouguin, T.; Hernandez, J.A.; Bull, J.; Pech, J.C.; Balague, C.
Mol. Gen. Genet. (1996) 251:81-90
Structure and expression of three genes encoding ACC oxidase homologs from melon (Cucumis melo L.).
#accession S66176
#status preliminary; nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-320 #label LAS
#cross-references EMBL:X95553; NID:g2104812; PID:e221407; PID:g1183900
#note the nucleotide sequence was submitted to the EMBL Data Library, February 1995

GENETICS
#introns 36/3; 223/3
CLASSIFICATION #superfamily 1-aminocyclopropane-1-carboxylate oxidase
SUMMARY #length 320 #molecular-weight 36397 #checksum 6287

Query Match 47.3%; Score 62; DB 2; Length 320;
Best Local Similarity 29.4%; Pred. No. 1.06e+01;
Matches 5; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

Db 282 FDDYMKLYGLKFOAKE 298
:|||||
QY 1 FSEFMROYVLISMOSQ 17

RESULT 13
ENTRY E70303 #type complete
TITLE hypothetical protein ag_040 - Aquifex aeolicus
ORGANISM #formal_name Aquifex aeolicus
DATE 08-May-1998 #sequence_revision 08-May-1998 #text_change 08-May-1998
E70303
A70300
Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Overbeek, R.; Sneed, M.A.; Keller, M.; Anjey, M.; Huber, R.; Feldman, R.A.; Short, J.M.; Olson, G.J.; Swanson, R.V.
Nature (1998) 392:353-358
The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
#cross-references MVID:98196666
#accession E70303
#status preliminary; nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-406 #label AOF
#cross-references GB:AE000670; NID:g2982779; PID:g2982788; GB:AE000657
#experimental_source strain VFS

GENETICS
#gene ag_040
SUMMARY #length 406 #molecular-weight 47559 #checksum 1115

Query Match 47.3%; Score 62; DB 2; Length 406;
Best Local Similarity 63.6%; Pred. No. 1.06e+01;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

DB 191 FEEFREFYVL 201
QY 1 FSEFMROYLV 11

RESULT 14
ENTRY 1
TITLE 1-muinnocyclopropane-1-carboxylic acid oxidase (EC 1.-.-.-) - muskmelon

ALTERNATE_NAMES ACC oxidase; ethylene-forming enzyme
ORGANISM #formal_name Cucumis melo #common_name muskmelon
DATE 31-Dec-1996 #sequence_revision 31-Dec-1996 #text_change 13-Sep-1998

ACCESSIONS JG6059; S29395; S66174
REFERENCE S66174

#journal Mol. Gen. Genet. (1996) 251:81-90
#title J.C.; Balaque, C.
Structure and expression of three genes encoding ACC oxidase homologs from melon (Cucumis melo L.).

#accession JG6059
#molecule_type mRNA
#residues 1-318 #label LAS
#cross-references EMBL:X95551; NID:g2104811; PID:e221404; PID:g1183896
#experimental_source seed, clone AC01

REFERENCE S29395
#authors Balaque, C.; Watson, C.F.; Turner, A.J.; Rouge, P.; Picton, S.; Pech, J.C.; Grierson, D.
Eur. J. Biochem. (1993) 212:27-34
Isolation of a ripening and wound-induced cDNA from Cucumis melo L. encoding a protein with homology to the ethylene-forming enzyme.

#accession S29395
#molecule_type mRNA
#status Preliminary
#residues 1-318 #label BAL
#cross-references EMBL:X69935; NID:g22662; PID:g22663

COMMENT This enzyme is a rate-limiting enzyme in ethylene biosynthesis. It catalyzes the last step of ethylene biosynthesis, converting 1-aminocyclopropane-1-carboxylic acid to ethylene, and regulates the rate of ethylene production.

GENETICS
#gene cm-ac01
#introns 36/2; 111/2; 222/3
CLASSIFICATION #superfamily 1-aminocyclopropane-1-carboxylate oxidase
KEYWORDS oxidoreductase; seed; transmembrane protein
FEATURE
278-298
39,177,234
SUMMARY #domain transmembrane #status predicted #label TMN
#binding_site Iron (His) #status predicted
#length 318 #molecular_weight 36127 #checksum 7777

Query Match 46.6%; Score 61; DB 2; Length 318;
Best Local Similarity 29.4%; Pred. No. 1.50e+01;
Matches 5; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

DB 284 FEDYMKLYGVFOAKE 300
QY 1 FSEFMROYLVLSMOSSQ 17

RESULT 15
ENTRY 1
TITLE S56767, #type complete
ORGANISM CDc46 homolog - mouse
DATE 10-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 04-Sep-1998

ACCESSIONS S56767
REFERENCE S56766

#authors Kimura, H.; Takizawa, N.; Nozaki, N.; Sugimoto, K.
#journal Nucleic Acids Res. (1995) 23:2097-2104
#title Molecular cloning of cDNA encoding mouse Cdc21 and Cdc46 homologs and characterization of the products: physical interaction between Pl(MCM3) and Cdc46 proteins.

#accession S56767
#status preliminary; nucleic acid sequence not shown; translation not shown

#molecule_type mRNA
#residues 1-733 #label KIM
#cross-references EMBL:D26090; NID:g940403; PID:d1005624; PID:g940404
#note the nucleotide sequence was submitted to the EMBL Data Library, December 1993

CLASSIFICATION #superfamily cell division control protein CDc47
FEATURE
327-539
SUMMARY #domain MCM2 core domain similarity #label MCM
#length 733 #molecular_weight 82342 #checksum 6080

Query Match 46.6%; Score 61; DB 2; Length 733;
Best Local Similarity 70.0%; Pred. No. 1.50e+01;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

DB 35 FKEFLROYRV 44
QY 1 FSEFMROYLV 10

Search completed: Fri Apr 16 14:22:14 1999
Job time : 18 secs.

THIS PAGE BLANK (USPTO)

TITLE	N23K, a gene transiently up-regulated during neural differentiation, encodes a precursor protein for a newly identified neuropeptide nociceptin
JOURNAL	Biochem. Biophys. Res. Commun. 217 (2), 539-545 (1995)
MEDLINE	96106851
REFERENCE	2 (bases 1 to 1047)
AUTHORS	Saito, Y.
TITLE	Direct Submission
JOURNAL	Submitted (05-MAR-1995) to the DDBJ/EMBL/GenBank databases. Yumiko Saito, Tokyo Metropolitan Institute for Medical Science, molecular biology; Honkomagome 3-18-22, Bunkyo-ku, Tokyo 113, Japan (E-mail:saito@rinsoken.or.jp, Tel:03-3823-2101(ex.5248), Fax:03-5565-6609)
COMMENT	On Mar 1, 1997 this sequence version replaced gi:1217905. Sequence updated (06-Mar-1996) by:Yumiko Saito.
FEATURES	Location/Qualifiers 1..1047 /organism="Mus musculus" /db_xref="taxon:10090" /cell_line="NS20Y" /cell_type="neuroblastoma" 57..620 /gene="N23K" 57..620 /gene="N23K" /gene="N23K" /note="neuropeptide precursor" /codon_start=1 /product="N23K protein" /db_xref="PID:d1009395" /db_xref="PID:g1217906" /translation="MKILFCVLLLSLVSFSSCPROCLTCQEKIHPAPSFNLKTG ILCEERYEPPLMTVTCTKVASSGGSLSPDPELVSAALQIPRASEKHUKRPVRVA SLVQRDAEPADAPGADAEPADEVEQKOLQRFEGFTGARSRARKLANQKRRE SEFRROYLPESMOSORRTLHONGNV"
CDS	polyA_signal 1040..1045 BASE COUNT 237 a 315 c 263 g 232 t ORIGIN
Query Match	98.0% Score 196 DB 29 Length 1047
Best Local Similarity	56.9% Pred. No. 5,94e-08;
Matches	29; Conservative 11; Mismatches 11; Indels 0; Gaps 0;
Db	477 TTGGGGCCTTACCCGCCCGGAATCAAGCCCCGGAAGTTGTGCACACCAG 527
Oy	1 TTYGGNGNTTYACGNCNGCMGNAMARSNGCMGAARYTGCMAYCAR 51
RESULT	10
LOCUS	HSHU48263 1198 bp mRNA PRI 23-AUG-1996
DEFINITION	Human pre-pro-orphanin FQ (OFO) mRNA, complete cds.
ACCESSION	U48263
NID	91185009
KEYWORDS	.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrate; Eutheria; Primates; Catarrhini; Hominoidea; Homo. 1 (bases 1 to 1198)
AUTHORS	Nothacker,H.P., Reinscheid,R.K., Mansour,A., Henningsen,R.A., Ardati,A., Monsma,F.J., Jr., Watson,S.J. and Clivelli,O.
TITLE	Primary structure and tissue distribution of the orphanin FQ precursor
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 93 (16), 8677-8682 (1996)
MEDLINE	96323383
REFERENCE	2 (bases 1 to 1198)
AUTHORS	Nothacker,H.-P. and Henningsen,R.A.
TITLE	Direct Submission
JOURNAL	Submitted (02-FEB-1996) Hans-Peter Nothacker, PRPN 69/202, Hoffmann-La Roche AG, Grenzacherstr 124, Basel, 4070, Switzerland
FEATURES	Location/Qualifiers 1..1198 /organism="Homo sapiens" /db_xref="taxon:9606"

FEATURES	Source	Location/Qualifiers
5'UTR	gene	/issue_type="brain"
gene		1..211
		212..742
		/gene="oFQ"
		212..742
		/gene="oFQ"
		/codon_start=1
		/product="pre-pro-orphanin FQ"
		/db_xref="PID:g1185010"
		/translation="MKVLCDLLILSLFSSVSSSCQDCVLCREKLPITLDSFLKWC ILDEEEVFPSPMTPTCTKYMARSWSLSPAAPEHYAALYOPASMSOHLRMPFVVR SLROEDPEPPEGMEEAGEEMQKQKRFGEFTGARKSARLANQKRSEEMROVLVLS MOSSQRRRTIHOXNGN"
		212..268
		/gene="oFQ"
		599..649
		/gene="oFQ"
		/product="orphanin FQ"
		743..1198
		1198
		/note="29 A nucleotides"
BASE COUNT	244 a	335 c 345 g 273 t 1 others
ORIGIN		
Query Match		98.0%; Score 196; DB 28; Length 1198;
Best Local Similarity		56.9%; Pred. No. 5,946-08;
Matches	29; Conservative	11; Mismatches 11; Indels 0; Gaps 0;
Db	599	TTTGGGGGCTTACCGGGGCGGGAAGTGGCCGAGGAGTGGCCATCAG 649
Oy	1	TTTGGGNGTITTYACNGGNCNGNNAHNSCNGCNGNAARTNGCNATCAR 51
RESULT	11	
LOCUS	AB005251	1209 bp mRNA MAM 09-APR-1998
DEFINITION	Bos taurus mRNA for nociceptin/orphanin FQ precursor, complete cds.	
ACCESSION	AB005251	
NID	g3041768	
KEYWORDS	nociceptin/orphanin FQ precursor.	
SOURCE	Bos taurus cDNA to mRNA.	
ORGANISM	Bos taurus	
REFERENCE	Eukaryote; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Artiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.	
AUTHORS	1 (sites)	
TITLE	Okuda-Ashitaka, E., Minami, T., Tachibana, S., Yoshihara, Y., Nishitani, Y., Kimura, T. and Ito, S. Nocistatin, a peptide that blocks nociceptin action in pain transmission	
JOURNAL	Submitted (26-JUN-1997) to the DDBJ/EMBL/GenBank databases. Emiko Nature 392 (6673), 286-289 (1998)	
MEDLINE	98180619	
REFERENCE	2 (bases 1 to 1209)	
AUTHORS	Okuda-Ashitaka, E.	
TITLE	Direct Submission	
JOURNAL	Submitted (26-JUN-1997) to the DDBJ/EMBL/GenBank databases. Emiko Okuda-Ashitaka, Kansai Medical University, Department of Medical Chemistry; 10-15 Fumizono, Moriyasu, Osaka, JAPAN, Moriyasu, Osaka 570, Japan (E-mail: ashitaka@takii.kmu.ac.jp, Tel: 06-992-1001, Fax: 06-992-1781)	
FEATURES		
Source	Location/Qualifiers	
	1..1209	
	/organism="Bos taurus"	
	/db_xref="taxon:9913"	
	210..740	
	/codon_start=1	
	/product="nociceptin/orphanin FQ precursor"	
	/db_xref="PID:d1026347"	
	/db_xref="PID:g3041769"	
	/translation="MKILFCDLLILSLFSSVSSSCQDCVLCREKLPITLDSFLKWC ILDEEEVFPSPMTPTCTKYMARSWSLSPADPHVAALDOPASMSOHLRMPFVVR SLROEDPEPPEGMEEAGEEMQKQKRFGEFTGARKSARLANQKRSEEMROVLVLS MOSSQRRRTIHOXNGN"	
BASE COUNT	240 a	359 c 341 g 269 t
ORIGIN		

REFERENCE	AUTHORS	JOURNAL	TITLE
1	Submitted (23-DEC-1995)	to the DBJ/EMBL/Genbank databases. Hitoshi Takeshima, University of Tokyo, Dept. of Pharmacology, Faculty of Medicine, Honjo 7-3-1, Bunkyo-ku, Tokyo 113, Japan (E-mail: takeshimem.u-tokyo.ac.jp, tel: 03-3812-2111 (ex. 3422), Fax: 03-3815-9360)	
2	(bases 1 to 1354)		
3	Unpublished (1996)		
4	Houtani, T., Nishi, M., Takeshima, H., Nukada, T. and Sugimoto, T.		
5	Structure and regional distribution of nociceptin/orphanin FQ precursor		
6	Biochem. Biophys. Res. Commun. 219 (3), 714-719 (1996)		
7	Location/Qualifiers		
8	1. 1354		
9	/organism="Mus musculus"		
10	/db_xref="taxon:10090"		
11	/dev_stage="adult"		
12	/tissue_type="brain"		
13	360..923		
14	/codon_start=1		
15	/product="nociceptin/orphanin FQ"		
16	/db_xref="PID:d1012281"		
17	/db_xref="PID:g1311473"		
18	/translation="MKILFCDVLLSLSSVFSSCPDCLTCOEKLPADSPFNLTLC ILCEEKVEFRPLMTVCTKVMASGSLSPADEIVSALAYORASMOHLKMPVR SLVORAEAGADAEAGADAEADAEADAEVQQLQKRFGGFIGARSAKLINQRF SEMRQIVLSMOSQRRRLHONGNV"		
19	polyA.signal		
20	BASE COUNT 309 a 370 c 386 g 289 t		
21	Query Match		
22	Best Local Similarity 56.9% Pred. NO. 5.94e-08;		
23	Matches 29; Conservative 11; Mismatches 11; Indels 0; Gaps 0;		
24	Db 780 TTGGGGGCTTCACCGGGCCCGGAATACACCGCGAATTGGCCAAACG 830		
25	QY 1 TTTGNGNGNTTYACNGCNGMNAHMSNGMNAHYTNGCNAAYCAR 51		
26	RESULT 14		
27	LOCUS HPAB000638 9956 bp DNA BCT 06-AUG-1997		
28	DEFINITION Helicobacter pylori section 116 of 134 of the complete genome.		
29	ACCESSION AE000638		
30	NID 92314547		
31	KEYWORDS Helicobacter pylori.		
32	SOURCE Helicobacter pylori.		
33	ORGANISM Helicobacter pylori		
34	REFERENCE 1 (bases 1 to 9956)		
35	1. Tomb, J.-F., White, O., Kerlavage, A.R., Clayton, R.A., Sutton, G.G., Fleischmann, R.D., Ketchum, K.A., Klenk, H.P., Gill, S., Dougherty, B.A., Nelson, K., Quackenbush, J., Zhou, L., Kirkness, E.F., Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khakh, H.G., Giodak, A., McKenney, K., Fitzgerald, L.M., Lee, N., Adams, M.D., Hickey, E.K., Berg, D.E., Gocayne, J.D., Uitterback, T.R., Peterson, J.D., Kelley, J.M., Cotton, M.D., Meldman, J.M., Fujii, C., Bowman, C., Wathey, L., Wallin, E., Hayes, W.S., Borodovsky, M., Karp, P.D., Smith, H.O., Fraser, C.M. and Venter, J.C.		
36	The complete genome sequence of the gastric pathogen Helicobacter pylori		
37	Nature 388 (6642), 539-547 (1997)		
38	97394467		
39	2 (bases 1 to 9956)		
40	1. Tomb, J.-F., White, O., Kerlavage, A.R., Clayton, R.A., Sutton, G.G., Fleischmann, R.D., Ketchum, K.A., Klenk, H.P., Gill, S., Dougherty, B.A., Nelson, K., Quackenbush, J., Zhou, L., Kirkness, E.F., Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khakh, H.G., Giodak, A., McKenney, K., Fitzgerald, L.M., Lee, N., Adams, M.D., Hickey, E.K., Berg, D.E., Gocayne, J.D., Uitterback, T.R., Peterson, J.D., Kelley, J.M., Cotton, M.D., Meldman, J.M., Fujii, C., Bowman, C., Wathey, L., Wallin, E., Hayes, W.S., Borodovsky, M., Karp, P.D., Smith, H.O., Fraser, C.M. and Venter, J.C.		
41	The complete genome sequence of the gastric pathogen Helicobacter pylori		
42	Nature 388 (6642), 539-547 (1997)		
43	97394467		
44	2 (bases 1 to 9956)		
45	1. Tomb, J.-F., White, O., Kerlavage, A.R., Clayton, R.A., Sutton, G.G., Fleischmann, R.D., Ketchum, K.A., Klenk, H.P., Gill, S., Dougherty, B.A., Nelson, K., Quackenbush, J., Zhou, L., Kirkness, E.F., Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khakh, H.G., Giodak, A., McKenney, K., Fitzgerald, L.M., Lee, N., Adams, M.D., Hickey, E.K., Berg, D.E., Gocayne, J.D., Uitterback, T.R., Peterson, J.D., Kelley, J.M., Cotton, M.D., Meldman, J.M., Fujii, C., Bowman, C., Wathey, L., Wallin, E., Hayes, W.S., Borodovsky, M., Karp, P.D., Smith, H.O., Fraser, C.M. and Venter, J.C.		
46	The complete genome sequence of the gastric pathogen Helicobacter pylori		
47	Nature 388 (6642), 539-547 (1997)		
48	97394467		
49	2 (bases 1 to 9956)		
50	1. Tomb, J.-F., White, O., Kerlavage, A.R., Clayton, R.A., Sutton, G.G., Fleischmann, R.D., Ketchum, K.A., Klenk, H.P., Gill, S., Dougherty, B.A., Nelson, K., Quackenbush, J., Zhou, L., Kirkness, E.F., Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khakh, H.G., Giodak, A., McKenney, K., Fitzgerald, L.M., Lee, N., Adams, M.D., Hickey, E.K., Berg, D.E., Gocayne, J.D., Uitterback, T.R., Peterson, J.D., Kelley, J.M., Cotton, M.D., Meldman, J.M., Fujii, C., Bowman, C., Wathey, L., Wallin, E., Hayes, W.S., Borodovsky, M., Karp, P.D., Smith, H.O., Fraser, C.M. and Venter, J.C.		
51	The complete genome sequence of the gastric pathogen Helicobacter pylori		
52	Nature 388 (6642), 539-547 (1997)		
53	97394467		
54	2 (bases 1 to 9956)		
55	1. Tomb, J.-F., White, O., Kerlavage, A.R., Clayton, R.A., Sutton, G.G., Fleischmann, R.D., Ketchum, K.A., Klenk, H.P., Gill, S., Dougherty, B.A., Nelson, K., Quackenbush, J., Zhou, L., Kirkness, E.F., Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khakh, H.G., Giodak, A., McKenney, K., Fitzgerald, L.M., Lee, N., Adams, M.D., Hickey, E.K., Berg, D.E., Gocayne, J.D., Uitterback, T.R., Peterson, J.D., Kelley, J.M., Cotton, M.D., Meldman, J.M., Fujii, C., Bowman, C., Wathey, L., Wallin, E., Hayes, W.S., Borodovsky, M., Karp, P.D., Smith, H.O., Fraser, C.M. and Venter, J.C.		
56	The complete genome sequence of the gastric pathogen Helicobacter pylori		
57	Nature 388 (6642), 539-547 (1997)		
58	97394467		
59	2 (bases 1 to 9956)		
60	1. Tomb, J.-F., White, O., Kerlavage, A.R., Clayton, R.A., Sutton, G.G., Fleischmann, R.D., Ketchum, K.A., Klenk, H.P., Gill, S., Dougherty, B.A., Nelson, K., Quackenbush, J., Zhou, L., Kirkness, E.F., Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khakh, H.G., Giodak, A., McKenney, K., Fitzgerald, L.M., Lee, N., Adams, M.D., Hickey, E.K., Berg, D.E., Gocayne, J		

TITLE
JOURNAL
COMMENT

Bowman, C., Matthey, L., Wallin, E., Hayes, W. S., Borodovsky, M., Karp, P. D., Smith, H. O., Fraser, C. M. and Venter, J. C.
 Direct Submission
 Submitted (06-AUG-1997) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA
 Address all correspondence to:

hpdb@tigr.org

or

J. Craig Venter
 The Institute for Genomic Research
 9712 Medical Center Dr.
 Rockville, MD, 20850

Douglas E. Berg is with the Washington University St. Louis School of Medicine, St. Louis, MO 63110.

Eric Wallin is with Stockholm University Department of Biochemistry, Stockholm, Sweden

Hamilton O. Smith is with the Johns Hopkins University School of Medicine, Baltimore, MD, 21205.

Mark Borodovsky and William S. Hayes are with the School of Biology, Georgia Tech, Atlanta, GA, 30332.

Peter D. Karp is with SRI International, Artificial Intelligence Center, Menlo Park, CA, 94025.

All other authors are with The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD, 20850.

Coding potential analysis was performed using Genemark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332.

e-mail: mark@amber.gatech.edu

Submission and annotation: Owen White
 e-mail: owhite@tigr.org

Biological role information, putative identifications, sequence alignments, on-line name and sequence search capability are available at TIGR's World Wide Web site.
 (URL: <http://www.tigr.org/tdb/hmdb/hpdb/hpdb.html>).

Location/Qualifiers

FEATURES
source
 1..9956
 /organism="Helicobacter pylori"
 /strain="26695"
 /db_xref="taxon:210"
 complement(116..556)
 /gene="HP1377"
 complement(116..556)
 /gene="HP1377"
 /note="hypothetical protein; identified by Genemark; putative"
 /codon_start=1
 /transl_table=11
 /product="H. pylori predicted coding region HP1377"
 /db_xref="PID:g2314555"
 /translation="MSLRRFCFAFLKQKSPMIFDVKADILGFETIHKMLQKIDEIFLRNLSSENSVVSFTLVNPFALRKVEFEPYPLKILLEEGAKSVYANIMVVOPIELSTVNYLADLILNLDKQLMGQVLDNSNKYPHYLRNIIISHTHE"
 914..1576
 /gene="HP1378"
 914..1576
 /gene="HP1378"
 /note="similar to EGAD:40678 percent identity: 25.53; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11

gene
 CDS

/product="competence lipoprotein (comL)"
 /db_xref="PID:g2314548"
 /translation="WRLNFKFTLEFITAIVIGTGCAKKKKKDEYKPAIFWYOGILREILFANLETADNYSSLQSEHISLPVPPAMALGQAHKKKKEVYSLASFPEYIKRGTQNDYLTFLFLQSHYYAFKNHSDQEFISISYLSLEFIEKYNRSPRVEYMOIKRFLGQNELNRAIANVYKKRHKPEGVKRLNIDETLEKETPKRSHMPVYLIFDM"
 1618..4125
 /gene="HP1379"
 1618..4125
 /gene="HP1379"
 /note="similar to EGAD:37242 percent identity: 43.94; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="ATP-dependent protease (lon)"
 /db_xref="PID:g2314549"
 /translation="MTEDPKTLLVLEDPEFVPMIAPFIQNNASIKAVYAKNKSLVFIACQKDKNDENAPYIDVGYIGSYMEANPNRKYLLFNGIKRILPEAKE

gene
 CDS

gene

CDS

NEQFLQKQISFTLEYTDKENQIAIVEYKKEVITLVANVSLFPPDLKALQEDDPNRIADLIAALHLKRDQAYSLFANNNTBQRLDLIDIVIEKTKQLQKEIKSKYHOKMEQTKKEFLKQLOKIOKEIGTDKQDQEDLNQYOKLESIKPFLKEEAFKIKQIDRLSRTHADSDSATQNYIEMLDVFPQYQKALDIKHVEQDLKDRHSYLRPKERIVEYFATMQLLEMRKKRPEKDKTKTGTLICFYGPVGKTSIANSIKRDEPIYRIALGGLEDVNLGHRRTYIGSPGRYVQGLIIRAKKNNPVMYDEIDKDYRSVQDPASALEIIDPQNTAFRDHYANFSDLSQVIFATANNIDIPAPLRDMERFVSSTPNKEEJAKNYLLIPQELKHALKPSVEISHECLKLIETRYRAGVDRARQATATMKLVALKLEIDNPQKGRTRKKNKSESDQKSENQKSENFPCVSTPPNKLKYLERMVFIEDPIDEENKIGIYNGLAWTVPVGGDVLEKLVKIRGKGLKLTGSVYMESAIIASFVIVLNDNETLKVPIIPSETDAGKKKKYKAYNADLHLHPGATPKDGP

gene

CDS

gene

CDS

SAGIMASVMSIICDRATRSYVATGTGLTSGEVLPIGGLEKELIAFAKAGIKTALI PVKTERDDELPAEVRNENLIVAKNIAEVLKTL"
 4164..4961
 /gene="HP1380"
 4164..4961
 /gene="HP1380"
 /note="similar to EGAD:8704 percent identity: 30.22; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="prephenate dehydrogenase (tyrA)"
 /db_xref="PID:g2314550"
 /translation="MGSGSLGALQEWGFRKSYTGDNHNLAKKALLTGLVDEVCYGFELCEDVFLAIPVEGIIIGCLKMTSIRKSATITDUGAKQIINIRKSIRKNTAAHPMCDETFYKASVKGQYENALVTLCDLESGTQVQVIAKBIIGVAKRLKMSNEHDTAVAYLSHLPVHYSLAVANSVLKQNDPEMILSLAGGFRDMSRLSRSSPLMWDIFKQNRNVLEATIKCEKEIYQAKWAIENNDESLAEMMAQANKLQEFM"
 5013..5246
 /gene="HP1381"
 5013..5246
 /gene="HP1381"
 /note="hypothetical protein; identified by Genemark; putative"
 /codon_start=1
 /transl_table=11
 /product="H. pylori predicted coding region HP1381"
 /db_xref="PID:g2314556"
 /translation="MRGAKWGIIWIFENDYRLFHMDSMDLPSSNOQTTRIDYLRGSIIVFAIYIISHSRNSNAYKTLIASNKPVLSHLEI"
 complement(5305..5697)
 /gene="HP1382"
 complement(5305..5697)
 /gene="HP1382"
 /note="hypothetical protein; identified by Genemark; putative"
 /codon_start=1
 /transl_table=11
 /product="H. pylori predicted coding region HP1382"
 /db_xref="PID:g2314557"
 /translation="MTNITPEAKNINRHSVLLVERGRLARKYHQVVEELTIKQGYTFSSKNIAISGFWMYHDTSLDYSYNAKSECFYIPNDQNYPLQEMRBDCKEYERVKQOVVFNKKNELNELPKYINNAKKT"
 5666..6148

gene

gene

CDS

gene

CDS


```
/gene="HP1383"
/note="similar to EGAD:44312 percent identity: 38.05,
identified by sequence similarity substitution"
```

```

/product="restriction modification system S subunit"
/db.xref="PID:g2114551"
/translation="MFASGVILVILKELATNQGFSILPEKLNNEFLYLLLTAK
KILKLASGSTEPLVSKNRKRLILPLPNEQDIALANLSDSYLTNDALLKEE
SVKRALSEELLSQKRLKGNQMRVRLGDIALYINMPSKIAFLGCKRLNLSFAR
"
complement(6173..6379)
/gene="HP1384"
complement(6173..6379)
/gene="HP1384"
/note="hypothetical protein; identified by Genemark;
putative"
/codon_start=1
/translation_table=11
/product="H. pylori predicted coding region HP1384"
/db.xref="PID:g2134559"
/translation="MQNSVKRKLFEERENDALLKQACOEKQVTSCLKCEQYLNCK

```

Query Match

psic local similarity 46.5%; pred No. 2.18e+01;
Matches 20; Conservative 8; Mismatches 15; Indels 0; Gaps 0;

Db 7520 GCGAGCTTTTTCGAGAATTGTTGCTCCTTAAACCGCAA 7562

44 GCNARYTTNCKNGCNSWYTTNCKNGCNCNGTRANCNCRA 2

RESULT	15
--------	----

	DATE	BY
DEFINITION	Drosophila melanogaster, chromosome 2R, region 54E8-54F2, p1 clone	VI DEC 1996

ACCESSION AC004328 AC003873 AC003607 AC003874

KEYWORDS

ORGANISM

Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;

REFERENCE

TITLE	Author	Journal	Year
Sequencing of <i>Drosophila</i> chromosome 2R, region 54E8-54F2	W. J. Gehring	Unpublished (1998)	

AUTHORS

Chew, M., Doyle, C.M., Farfan, D.E., Flanagan, J., Houston, K.A.,

Lomotan, M.A., Mak, J., Mazda, P., Mok, M.S., Moshrefi, A.R.,

Snir, E., Twomey, B., Wan, K.H., Whitelaw, K.R., Yee, A., Zhang, R.,

TITLE	Direct Submission	Total
Sub-44-107-1000		

Laboratory, MS 64-121, Berkeley, CA 94720, USA

U
I
C
C
C
C
C

Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720

For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive web site (<http://www.trinityfly.org/sequence/>) or send email
to bdg@trinityfly.berkeley.edu.

P1 library location: 23-43.
Location/Qualifiers
1..55452

```

/organism="Drosophila melanogaster"
/strain="y2: cn bw sp"
/db_xref="taxon:7227"
/chromosome="2R"
/map="34E8-54F2"
/clone_lib="p1 library, partial Sauza in pad10sacB1"
/clone="p1 DS02155 (D221)"
/note="DS02155 (D221) is submitted as a bridge. It
minimally overlaps its proximal neighbor DS06848 (D168)
and its distal neighbor DS07328 (D169)."
BASE COUNT      15556 a 12344 c 12189 g 13363 t
ORIGIN

```

Query Match	55.5%;	Score 111;	DB 19;	Length 55452;
Best Local Similarity	41.7%;	Pred. No. 2.18e+01;		
Matches	20;	Conservative	10;	Mismatches 18;
			Indels	0;
			Gaps	0

D_b 19408 TTTGGTITTCGCACACTGGCCCCAAAAAATCCGCACGTAACTCTGCGAAT 19455
||:|| :||| : ||:: || : ||:: ||||
QY 1 TTYYGNGNTTYACNGGNCMGNNAARWSNCGCNMGNAARYTINGCNAAY 48

Search completed: Tue Apr 27 10:14:01 1999
Job time : 172 secs.

RESULT 2
ID W25219 standard; peptide; 17 AA.
AC W25219;
DT 12-JAN-1998 (first entry)
DE Rat orphanin FQ receptor-binding peptide inhibitor.
KW Orphanin FQ receptor; binding; locomotor disease; diagnosis;
treatment; opioid inhibitor; opiate induced hypothermia; drug design;
morphine induced analgesia; methadone specific opioid receptor; ss.
OS Synthetic.
PN W09707212-A1.
PD 27-FEB-1997.
PF 12-AUG-1996; U13305.
PR 03-NOV-1995; US-553058.
PR 11-AUG-1995; US-514451.
PA (UOOR)-UNIT OREGON HEALTH SCI.
PI Bunzow JR, Clivelli O, Grandy DK, Grisell JE, Mogil JS;
PI Monsuma FJ, Notchacker H-P, Reinscheid RK;
WP1: 97-165296/15.
PT Mammalian orphanin FQ receptor inhibitor - used in the diagnosis and
treatment of locomotor disease.
PS Claim 12: Page 48: 68pp: English.
CC W25219 and W25220 are peptide inhibitors of the rat orphanin FQ (OFQ)
receptor.
CC Highly specific peptides that bind the OFQ receptor were identified as
OFQ receptor inhibitors. The peptides can be used to antagonise a
physiological effect of an opioid in an animal. The peptides antagonise
opiate induced hypothermia and morphine induced analgesia in animals.
CC They may also be used in the diagnosis and treatment of locomotor
disease. The peptides may also be used in the design of a methadone
specific opioid receptor (MSOR), in drug design and for the isolation
of endogenous receptors for anti-opioid agonists and antagonists found
in plasma, serum, lymph, cerebrospinal fluid etc.
SQ Sequence 17 AA;

Query Match	100.0%;	Score 116;	DB 25;	Length 17;
Best Local Similarity	100.0%;	Pred. No. 6.78e-06;		
Matches	17;	Conservative 0;	Mismatches 0;	Indels 0;
			Gaps 0;	

```

Db      1  fggftgarksarklang 17
        |||||
Qy      1  FGFTGARKSARKLANQ 17

```

RESULT	3
ID	W25220 standard; peptide; 17 AA.

DT	12-JAN-1998	(first entry)
DE	rat orphanin FQ receptor-binding peptide inhibitor.	
KW	Orphanin FQ receptor; binding; locomotor disease; diagnosis;	
KW	treatment; opioid inhibitor; opiate induced hypothermia; drug design;	
KW	morphine induced analgesia; methadone specific opioid receptor; ss.	
OS	Synthetic.	
PN	WO9707212-A1.	
PD	27-FEB-1997.	
PD	12-AUG-1996; U13305.	
PR	03-NOV-1995; US-553058.	
PR	11-AUG-1995; US-514451.	
PT	(CYOR-) UNIV OREGON HEALTH SCI.	
PT	Bunzow JR, Civelli O, Grandy DK, Grisel JE, Mogil JS;	
PI	Monsma FJ, Natchacker H-P, Reinscheid RK;	
DR	WPI: 97-165296/15.	
PT	Mammalian orphanin FQ receptor inhibitor - used in the diagnosis and	
PT	treatment of locomotor disease.	
PS	Claim 12; Page 48; 68pp; English.	
CC	W23219 and W25220 are peptide inhibitors of the rat orphanin FQ (OFQ)	
CC	receptor.	
CC	Highly specific peptides that bind the OFQ receptor were identified as	
CC	OFQ receptor inhibitors. The peptides can be used to antagonise a	
CC	physiological effect of an opioid in an animal. The peptides antagonise	
CC	opiate induced hypothermia and morphine induced analgesia in animals.	
CC	They may also be used in the diagnosis and treatment of locomotor	
CC	disease. The peptides may also be used in the design of a methadone	

CC specific opioid receptor (MSOR), in drug design and for the isolation
CC of endogenous receptors for anti-opioid agonists and antagonists found
CC in plasma, serum, lymph, cerebrospinal fluid etc.
50 Sequence 17 AA;

Query Match	90.5%;	Score 105;	DB 25;	Length 17;
Best Local Similarity	94.1%;	Pred. No. 1.46e-04;		
Matches 16;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;

```
Db 1 fggftgarksarkyang 17
    |||||
Qy 1 FGGFTGARKSARKLANQ 17
```

RESULT	4
ID	R38779 standard; peptide; 113 AA

DT	22-DEC-1993 (first entry)
DE	Active site peptide fragment #18.
KW	Active site; active region; prediction; secondary structure

OS	Synthetic.
PN	J05155899-A.
PD	22-JUN-1993.

PR 23-AUG-1991; JP-212284.
PA (SAGA) SAGAMI CHEM RES CENTRE.
DR WPI; 93-232353/29.

PT polypeptide - by determ. of sec. structure energy values of
PT partial regions along prim. sequence of physiologically active
PT polynucleotide, etc.

CC The sequences given in R38/62-79 are peptides which were used to
CC demonstrate the method of the invention. These peptides represent
CC active sites/regions and the method of the invention may be used to

values of the secondary structure of various partial regions of the polypeptide to be tested and the primary sequence as well. At least one partial region should show a local peak energy value

CC This method may be used in the elucidation of the mechanism of
CC action of polypeptides or polynucleotides, and in the engineered
CC improvement of such sequences.

Query Match	56.9%;	Score 66;	DB 8;	Length 113;
Best Local Similarity	53.3%;	Pred. NO. 4.53e+00;		
Matches	8;	Conformation	1;	Total

Dd 13 g1sggrkpprk1vr 27
|::|::|::|::|:
Ou 3 CETGCAVRSABRIANO 17

TD	RESULT
P00000	5

AC	R90989;
DT	27-MAY-1996 (first entry)
DE	Human adrenergic G-protein coupled receptor.
KW	G-protein coupled receptor polypeptide; GPCR; adrenergic receptor.

KW agonistic; antagonistic; therapy; hypertension; respiratory disease
 OS Homo sapiens.
 PN W09605225-A1.
 PD 22-FEB-1996

PE 10-AUG-1994; 009051.
PR 10-AUG-1994; WO-U09051.
PA (HUMA-) HUMAN GENOME SCI INC
PI Adams MD 1 f y Sonnet DR.

DR WPI; 96-139642/14.
DR N-PSDB; T13002.
PT New isolated G-protein coupled receptor polypeptide - used to
PT develop nodes to inhibit or stimulate adrenergic receptors for

PT treating e.g. hypertension or respiratory disorders
 PS Claim 4; Fig 1; 60pp; English.
 CC Human G-protein coupled receptor polypeptide (GCRP) (R90989)
 CC is a protein structurally related to the alpha-1 adrenergic
 CC receptor family. It can be produced by expression of a cDNA clone
 CC (T13002) isolated from a human infant brain cDNA library.
 CC Recombinant GCRP is used to screen for agonist and antagonist
 CC cpds. useful e.g. in the treatment of hypertension and respiratory
 CC disorders, and to raise anti-GCRP antibodies.
 SQ Sequence 529 AA;

Query Match 50.9%; Score 59; DB 16; Length 529;
 Best Local Similarity 50.0%; Pred. No. 2.53e+01;
 Matches 8; Conservative 5; Mismatches 2; Indels 1; Gaps 1;

Db 497 gfgfrrgsrlvsq 511
 |||:::|:
 QY 2 GFTGARKSARKLANQ 17

RESULT 6
 ID R76600 standard: Protein; 370 AA.
 AC R76600;
 DT 29-FEB-1996 (first entry)
 DE Chromobacterium esterase.
 KW Esterase; mass production; genetic engineering; plasmid; expression;
 KM recombinant.
 OS Chromobacterium sp.
 PN J07163364-A.
 PD 27-JUN-1995.
 PF 15-DEC-1993; 315497.
 PR 15-DEC-1993; JP-315497.
 PA (SUMO) SUMITOMO CHEM CO LTD.
 DR WPI; 95-260046/34.
 DR N-PSDB; Q92918.
 PT Esterase gene - for mass production of esterase
 PS Claim 1; Page 12-14; 15pp; Japanese.
 CC The chromobacterium esterase gene has been isolated and sequenced.
 CC It is found to be 1639 base pairs (bp) long with a CDS of 1110 bp
 CC which codes for the 370 amino acid esterase protein. The gene
 CC can be used in recombinant systems to allow the mass prodn. of
 CC esterase in high yields.
 SQ Sequence 370 AA;

Query Match 49.1%; Score 57; DB 14; Length 370;
 Best Local Similarity 33.3%; Pred. No. 4.09e+01;
 Matches 5; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

Db 138 gfgfrrgsrlvdd 152
 |||:::|:
 QY 3 GFTGARKSARKLANQ 17

RESULT 7
 ID R86303 standard: protein; 370 AA.
 AC R86303;
 DT 16-APR-1996 (first entry)
 DE Thermostable esterase with mutation at position 160 and/or 189.
 KW Thermostable; esterase; mutant; organic synthesis; site specific.
 OS Chromobacterium sp.
 FH Key Location/Qualifiers
 FT misc-difference 160
 FT /note="site specific mutation, optional when
 FT wild type Gly189 is also mutated"
 FT misc-difference 189
 FT /note="site specific mutation, optional when
 FT wild type Gly160 is also mutated"
 PN J07213280-A.
 PD 15-AUG-1995.
 PF 02-FEB-1994; 011014.
 PR 02-FEB-1994; JP-011014.
 PA (SUMO) SUMITOMO CHEM CO LTD.
 DR WPI; 95-315922/41.

PT Thermostable esterase with mutation at amino acid position 160
 PT and/or 189 - useful for organic synthesis, also mutant gene and
 PT transformed microbe for recombinant prodn. of enzyme
 PS Claim 1; Pages 8-9; 16pp; Japanese.
 CC A thermostable Chromobacterium sp. esterase is produced when the
 CC wild type amino acid residues Gly160 and/or Gly189 undergo site
 CC specific mutation(s). The mutant thermostable esterase (R86303)
 CC can be used in organic synthesis.
 SQ Sequence 370 AA;

Query Match 49.1%; Score 57; DB 15; Length 370;
 Best Local Similarity 33.3%; Pred. No. 4.09e+01;
 Matches 5; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

Db 138 gfgfrrgsrlvdd 152
 |||:::|:
 QY 3 GFTGARKSARKLANQ 17

RESULT 8
 ID R29516 standard: Protein; 1385 AA.
 AC R29516;
 DT 20-APR-1993 (first entry)
 DE Bt toxin 17A.
 KW nematode worms; nematicide; nematicidal toxin; agriculture; plants;
 KM crops; pests; CryV proteins.
 OS Bacillus thuringiensis.
 PN BP-517367-A.
 PD 09-DEC-1992.
 PF 01-MAY-1992; 303969.
 PR 03-MAY-1991; US-693018.
 PR 31-JAN-1992; US-830050.
 PR 23-APR-1992; US-871510.
 PA (MYCO) MYCOGEN CORP.
 PI Fonceriada L, Narva KE, Payne JM, Schnepf HE, Schwab GE;
 DR WPI; 92-408829/50.
 DR N-PSDB; Q32159.
 PT Nematocidal toxins from Bacillus thuringiensis - useful for
 PT control of animal or plant parasites, deoxyribonucleic acid
 PT coding sequences, transformed hosts and transgenic plants
 PS Claim 1(b); Page 23; 57pp; English.
 CC This sequence represents the Bacillus thuringiensis delta-endotoxin
 CC 17A having nematicidal activity. It was decoded from the appropriate
 CC DNA. Intact cells expressing this toxin may be used in nematicidal
 CC compans, either sprayed on plants to kill agricultural crop pests,
 CC or taken as a drug to kill gastro-intestinal worms in mammalian hosts.
 SQ Sequence 1385 AA;

Query Match 48.3%; Score 56; DB 6; Length 1385;
 Best Local Similarity 50.0%; Pred. No. 5.19e+01;
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Db 943 fgekkalkrlvng 956
 |:::|:
 QY 4 FTGARKSARKLANQ 17

RESULT 9
 ID R28889 standard: Protein; 1385 AA.
 AC R28889;
 DT 01-APR-1993 (first entry)
 DE Toxin 17A.
 KW Endotoxin; acarides; pest; Two Spotted Spider; mite;
 KM phytophagus.
 OS Bacillus thuringiensis ps17.
 FH Key Location/Qualifiers
 FT peptide
 FT /note="amino acid sequence of a probe for cloning
 FT acaride-active genes"
 PN W09219106-A.
 PD 12-NOV-1992.
 PF 30-APR-1992; U03546.
 PR 30-APR-1991; US-6932210.

PR 13-SEP-1991: US-759248.
 PR 30-SEP-1991: US-768141.
 PA (MYCO) MYCOGEN CORP.
 PI Bagley AL, Cannon RJC, Payne JM;
 DR WPI: 92-398411/48.
 DR N-PSDB: Q20803.
 PT New Bacillus thuringiensis isolates and toxins - used for
 controlling acarid pests of livestock, fowl, stored prods. and
 plants
 PT Claim 18: Page 53 + 27-30; 62pp; English.
 CC Gene sequences encoding a toxin which is active against acarides and
 is obtainable from B. thuringiensis isolates PS17a, PS17b, 33f2,
 CC PS52a1, PS59d1, PS66a1 and PS50C are given in Q30803-07 and Q30820-21
 CC respectively. The toxin is a delta-endotoxin active against acarid
 CC pests, including the Two Spotted Spider mite. The isolates can be
 CC used against non-phytophagous mites such as acarid pests of livestock,
 CC fowl and stored prods. The genes can be cloned and used to
 CC transform other hosts, which can be used to control mites, or in
 CC the case of transgenic plants, be resistant to mites.
 SQ Sequence 1385 AA;
 Query Match 48.3%; Score 56; DB 6; Length 1385;
 Best Local Similarity 50.0%; Pred. No. 5.19e+01;
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
 Db 943 fgdekalkrlvng 956
 | : | | | | |
 QY 4 FTGARKSARKLANQ 17
 RESULT 10
 ID R58631 standard; Protein: 1385 AA.
 AC R58631;
 DE 16-MAY-1995 (first entry)
 DE Bacillus thuringiensis PS17a delta-endotoxin protein.
 DE Endotoxin; Bacillus thuringiensis; PS17; toxin; delta-endotoxin;
 KW parasporal; crystalline; inclusion; toxic; pest; spider mite;
 KW Tetranychus urticae; house dust mite; Dermatophagoides pteronyssinus;
 KW phytophagous; mite; acarid; livestock; fowl; acaridicidal; insect;
 KW virus; plant; control; transgenic; resistance.
 OS Bacillus thuringiensis strain PS17.
 PS US5350576-A.
 PD 27-SEP-1994.
 PF 13-SEP-1991: 759248.
 PR 13-SEP-1991: US-759248.
 PR 30-SEP-1991: US-768141.
 PR 17-MAY-1993: US-063170.
 PA (MYCO) MYCOGEN CORP.
 PI Cannon RJC, Payne J, Ralph AL;
 DR WPI: 94-310547/38.
 DR N-PSDB: Q72179.
 PT Controlling house dust mite - using new delta endotoxin isolates
 PT from Bacillus thuringiensis with acaridicidal activity
 PS Claim 5: Column 13-16; 10pp; English.
 CC The amino acid sequence of a novel delta-type endotoxin from Bacillus
 CC thuringiensis (B.t.) strain PS17. The toxin gene encodes a protein of
 CC 1385 amino acids. The PS17 delta-endotoxins range in mol. wt. from 128
 CC to 155 kDa. The delta-endotoxins are derived from parasporal crystalline
 CC protein inclusions and are highly toxic to pests and specific in their
 CC activity. The toxins presented (Q72179-80) are specific to the
 CC two-spotted spider mite (Tetranychus urticae) and to the house dust mite
 CC (Dermatophagoides pteronyssinus). The use of B.t. isolates containing
 CC the endotoxins can also target non-phytophagous mites such as acarid
 CC pests of livestock, fowl and stored products. The gene(s) from the B.t.
 CC isolates which encode the acaridicidal endotoxin can be cloned from the
 CC isolates and used to transform other hosts e.g. prokaryotic cells, insect
 CC viruses, eukaryotic cells or plants, which could be used to control mites
 CC or, in the case of transgenic plants, confer resistance to mites.
 SQ Sequence 1385 AA;
 Query Match 48.3%; Score 56; DB 12; Length 1385;
 Best Local Similarity 50.0%; Pred. No. 5.19e+01;
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Db 943 fgdekalkrlvng 956
 | : | | | | |
 QY 4 FTGARKSARKLANQ 17
 RESULT 11
 ID W13884 standard; Protein: 1385 AA.
 AC W13884;
 DE 14-MAY-1997 (first entry)
 DE 17a toxin.
 KW Toxin; ant; Bacillus thuringiensis; hymenopteran pest; pharaoh ant;
 KW biological control; Monomorium pharaonis; delta-endotoxin; lepidoptera;
 KW insect.
 OS Bacillus thuringiensis isolate PS17a.
 PS US5596071-A.
 PD 21-JAN-1997.
 PF 22-MAY-1991: 703977.
 PR 22-MAY-1991: US-703977.
 PR 25-NOV-1991: US-797645.
 PR 22-MAY-1992: US-887980.
 PR 24-NOV-1993: US-158232.
 PA (MYCO) MYCOGEN CORP.
 PI Foncerrada L, Fu J, Kennedy MK, Meier H, Payne JM;
 PI Randall JB, Schnepf HE, Schwab GE, Vick HJ;
 DR WPI: 97-107615/10.
 DR N-PSDB: T60069.
 PT Bacillus thuringiensis toxin - active against hymenopteran pests
 PS Disclosure: Column 35-42; 64pp; English.
 CC This sequence represents the 17a toxin isolated from the Bacillus
 CC thuringiensis (B.t.) isolate PS17a. B.t. is a gram-positive, spore
 CC forming, soil bacterium, characterised by parasporal crystalline protein
 CC inclusions. These proteins can be highly toxic to pests, and have been
 CC used to produce insect resistant plants. The previously isolated B.t.
 CC delta-endotoxins were mainly active against lepidopteran insects, however
 CC the proteins of the invention are active against hymenopteran insects.
 CC This sequence is an example of a toxin of the invention, for which the
 CC sequences shown in W13888 and W13871 represent the generic formulae. As
 CC the toxins of the invention are active against hymenopteran pests, they
 CC can be used for the biological control of ants, particularly pharaoh ants
 CC (Monomorium pharaonis).
 SQ Sequence 1385 AA;
 Query Match 48.3%; Score 56; DB 21; Length 1385;
 Best Local Similarity 50.0%; Pred. No. 5.19e+01;
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
 Db 943 fgdekalkrlvng 956
 | : | | | | |
 QY 4 FTGARKSARKLANQ 17
 RESULT 12
 ID R28803 standard; Protein: 1385 AA.
 AC R28803;
 DE 30-MAR-1993 (first entry)
 DE Bt toxin 17a.
 KW nematode worms; nematocidal; nematocidal toxin; agriculture; plants;
 KW crops; pests; CYYV proteins.
 OS Bacillus Thuringiensis.
 PS W09219739-A.
 PD 12-NOV-1992.
 PF 01-MAY-1992: U03624.
 PR 03-MAY-1991: US-693018.
 PR 31-JAN-1992: US-830050.
 PR 23-APR-1992: US-871510.
 PA (MYCO) MYCOGEN CORP.
 PI Foncerrada L, Narva KE, Payne JM, Schnepf HE, Schwab GE;
 DR WPI: 92-39866/48.
 DR N-PSDB: Q30935.
 PT New genes and toxins against nematodes - obtd. from Bacillus
 PT Thuringiensis isolates with nematocidal activity
 PS Claim 1(b); Page 42; 77pp; English.

CC This sequence represents the *Bacillus thuringiensis* delta-endotoxin
 CC 17a having nematocidal activity. It was decoded from the appropriate
 CC DNA. Intact cells expressing this toxin may be used in nematocidal
 CC compounds, either sprayed on plants to kill agricultural crop pests,
 CC or taken as a drug to kill gastro-intestinal worms in mammalian hosts.
 SQ Sequence 1385 AA;

Query Match 48.3%; Score 56; DB 6; Length 1385;
 Best Local Similarity 50.0%; Pred. No. 5.19e+01;
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Db 943 fgdekaklrkivng 956
 | : | | | | |
 QY 4 FTGARKSARKLANQ 17

RESULT 13
 ID R29026 standard; Protein; 1385 AA.
 AC R29026;
 DT 20-APR-1993 (first entry)
 DE *Bacillus thuringiensis* toxin protein 17a.
 KW Toxin protein; ant.
 OS *Bacillus thuringiensis* strain PS17.
 PN MO9220802-A.
 PD 26-NOV-1992.
 PF 22-MAY-1992; U04316.
 PR 22-MAY-1991; US-703977.
 PR 25-NOV-1991; US-797645.
 PR 12-MAY-1992; EP-304228.
 PA (MYCO) MYCOGEN CORP.
 PI Kennedy MK, Meier H, Payne JM, Randall JB, Uick HU;
 DR WPI: 92-415780/50.
 DR N-PSDB; 031409.
 PT Toxin proteins isolated from *Bacillus thuringiensis* - for controlling
 PT ants. e.g. fire, carpenter, Argentine and pharaoh ants
 PS Disclosure: Page 35; 71pp; English.
 CC *Bacillus thuringiensis* toxin protein 17a is useful as a method of
 CC biological control of ants, e.g. fire ants, carpenter ants,
 CC Argentine ants and pharaoh ants, as an alternative to chemical
 CC insecticides.
 SQ Sequence 1385 AA;

Query Match 48.3%; Score 56; DB 6; Length 1385;
 Best Local Similarity 50.0%; Pred. No. 5.19e+01;
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Db 943 fgdekaklrkivng 956
 | : | | | | |
 QY 4 FTGARKSARKLANQ 17

RESULT 14
 ID R76112 standard; Protein; 1385 AA.
 AC R76112;
 DT 27-FEB-1996 (first entry)
 DE *Ps17a* acaricide-active toxin from strain PS17.
 KW Insecticide; toxin; acaricide-active toxin; delta-endotoxin; PS17a; PS17;
 KW two spotted spider mite; tetranychus urticae; *Bacillus thuringiensis*.
 OS *Bacillus thuringiensis* strain PS17.
 PN Key location/Qualifiers
 FT region 2..15
 FT US5424410-A.
 PD 13-JUN-1995.
 PF 30-APR-1991; 693210.
 PR 30-APR-1991; US-693210.
 PR 13-SEP-1991; US-759248.
 PR 30-SEP-1991; US-768141.
 PR 30-APR-1992; US-876280.
 PR 03-NOV-1993; US-147188.
 PA (MYCO) MYCOGEN CORP.
 PI Bagley AL, Cannon RJC, Payne JW;
 DR WPI: 95-223727/29.

DR N-PSDB; 094053.
 PT DNA encoding delta endotoxin of *Bacillus thuringiensis* - used for
 PT the control of acaride pests and for the production of acaride pest
 PT resistant plants
 PS Example 3; Column 27-34; 42pp; English.
 CC The sequences represented by R76112 and R76113 represent acaride-active
 CC toxins isolated from *B. thuringiensis* isolate PS17. This sequence
 CC represents PS17a. These sequences are delta-endotoxins.
 CC Delta-endotoxins are insect toxins which act on insects in the orders
 CC Lepidoptera and Diptera. The toxin acts to destroy insect gut epithelium
 CC cells. The toxin is ingested in crystalline form, by larvae, and is
 CC transformed into biologically active forms by the insect gut juice
 CC proteases (in susceptible insects). The N-terminal region of this
 CC sequence was used in the design of probes for this sequence and other
 CC acaride-active toxins. The DNA encoding this sequence can be inserted
 CC into a vector which can be introduced into plant cells, thereby
 CC conferring insect resistance. This protein can be combined with other
 CC insecticidal proteins to increase the spectrum of activity and to give
 CC complete control of target pests. Plants transformed with the DNA
 CC encoding this sequence become resistant to acaride pests, e.g. the two
 CC spotted spider mite, *Tetranychus urticae*.
 SQ Sequence 1385 AA;

Query Match 48.3%; Score 56; DB 14; Length 1385;
 Best Local Similarity 50.0%; Pred. No. 5.19e+01;
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Db 943 fgdekaklrkivng 956
 | : | | | | |
 QY 4 FTGARKSARKLANQ 17

RESULT 15
 ID R20066 standard; Protein; 1385 AA.
 AC R20066;
 DT 26-MAR-1992 (first entry)
 DE *B. thuringiensis* toxin PS17a.
 KW worm; nematocide; flukicide; anthelmintic; parasite.
 OS *Bacillus thuringiensis*.
 PN EP-462721-A.
 PD 27-DEC-1991.
 PF 04-JUN-1991; 305047.
 PR 11-JUN-1990; US-535810.
 PR 24-JUL-1990; US-557246.
 PR 27-JUL-1990; US-558738.
 PR 10-AUG-1990; US-565544.
 PR 14-MAR-1991; US-669126.
 PR 27-MAR-1991; US-675772.
 PR 03-MAR-1991; US-695018.
 PA (MYCO-) MYCOGEN CORP.
 PI Narva KE, Payne JM, Schwab GE, Hickie LA, Galasan T;
 PI Sick AJ;
 DR WPI: 92-001086/01.
 DR N-PSDB; 020334.
 PT New *Bacillus thuringiensis* strains expressing toxins - have
 PT nematocidal activity, to control nematodes, helminths and flukes
 PT e.g. liver fluke *Fasciola hepatica*
 PS Claim 6; Page 24; 47pp; English.
 CC The PS17a toxin gene was isolated from B.t. strain PS17. Probe
 CC Q20333 was used to screen a PS17 total cellular DNA library. The
 CC gene was subcloned in a vector to give plasmid pYK2311 which was
 CC transformed into the acrycristalliferous B.t. strain, HD-1 cryB. The
 CC recombinant B.t. strain was grown to sporulation and the toxin
 CC protein purified by NaBr gradient centrifugation.
 CC See also Q20335-Q20343.
 SQ Sequence 1385 AA;

Query Match 48.3%; Score 56; DB 3; Length 1385;
 Best Local Similarity 50.0%; Pred. No. 5.19e+01;
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Db 943 fgdekaklrkivng 956
 | : | | | | |

Wed Apr 28 09:08:31 1999

US-09-011-797-2.rag

Page 6

OY 4 FTGAKKSARKLANQ 17

Search completed: Fri Apr 16 14:17:41 1999
Job time : 17 secs.

 M P E R E I T
 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Mparch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Fri Apr 16 14:19:10 1999; Maspar time 4.96 Seconds
 170.573 Million cell updates/sec
 Tabular output not generated.

Title: >US-09-011-797-2
 Description: (1-17) from US09011797.pep
 Perfect Score: 116
 Sequence: 1 FGFTGARKSARKLANQ 17

Scoring table:
 PAM 150
 Gap 15

Searched: 165420 segs, 49795644 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database:

sptrembl6
 1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
 5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organelle
 9:sp-phage 10:sp-plant 11:sp-rodent 12:sp-unclassified
 13:sp-vertebrate 14:sp-virus

Statistics: Mean 26.722; Variance 32.797; scale 0.815

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	116	100.0	176	6	062647	8.09e-15
2	66	56.9	526	1	P95887	4.09e-02
3	60	51.7	776	4	060719	7.95e-01
4	59	50.9	177	1	028463	1.28e+00
5	59	50.9	211	2	066484	1.28e+00
6	57	49.1	133	4	099899	3.27e+00
7	57	49.1	152	10	041349	3.27e+00
8	57	49.1	614	4	043245	3.27e+00
9	57	49.1	801	4	013436	3.27e+00
10	57	49.1	1061	10	004987	3.27e+00
11	56	48.3	890	10	P93024	5.18e+00
12	56	48.3	902	10	064965	5.18e+00
13	56	48.3	1385	2	045760	5.18e+00
14	55	47.4	470	10	040346	8.16e+00
15	55	47.4	476	10	040347	8.16e+00
16	54	46.6	224	1	027716	1.28e+01
17	54	46.6	316	2	038772	1.28e+01
18	54	46.6	378	5	016903	1.28e+01
19	54	46.6	593	10	P93002	1.28e+01
20	54	46.6	651	2	069635	1.28e+01

Result ID	Score	Query Match	Length	DB ID	Description	Pred. No.
1	116	100.0	176	6	062647	8.09e-15
2	66	56.9	526	1	P95887	4.09e-02
3	60	51.7	776	4	060719	7.95e-01
4	59	50.9	177	1	028463	1.28e+00
5	59	50.9	211	2	066484	1.28e+00
6	57	49.1	133	4	099899	3.27e+00
7	57	49.1	152	10	041349	3.27e+00
8	57	49.1	614	4	043245	3.27e+00
9	57	49.1	801	4	013436	3.27e+00
10	57	49.1	1061	10	004987	3.27e+00
11	56	48.3	890	10	P93024	5.18e+00
12	56	48.3	902	10	064965	5.18e+00
13	56	48.3	1385	2	045760	5.18e+00
14	55	47.4	470	10	040346	8.16e+00
15	55	47.4	476	10	040347	8.16e+00
16	54	46.6	224	1	027716	1.28e+01
17	54	46.6	316	2	038772	1.28e+01
18	54	46.6	378	5	016903	1.28e+01
19	54	46.6	593	10	P93002	1.28e+01
20	54	46.6	651	2	069635	1.28e+01

ALIGNMENTS

Result ID	Score	Query Match	Length	DB ID	Description	Pred. No.
1	116	100.0	176	6	062647	8.09e-15
2	66	56.9	526	1	P95887	4.09e-02
3	60	51.7	776	4	060719	7.95e-01
4	59	50.9	177	1	028463	1.28e+00
5	59	50.9	211	2	066484	1.28e+00
6	57	49.1	133	4	099899	3.27e+00
7	57	49.1	152	10	041349	3.27e+00
8	57	49.1	614	4	043245	3.27e+00
9	57	49.1	801	4	013436	3.27e+00
10	57	49.1	1061	10	004987	3.27e+00
11	56	48.3	890	10	P93024	5.18e+00
12	56	48.3	902	10	064965	5.18e+00
13	56	48.3	1385	2	045760	5.18e+00
14	55	47.4	470	10	040346	8.16e+00
15	55	47.4	476	10	040347	8.16e+00
16	54	46.6	224	1	027716	1.28e+01
17	54	46.6	316	2	038772	1.28e+01
18	54	46.6	378	5	016903	1.28e+01
19	54	46.6	593	10	P93002	1.28e+01
20	54	46.6	651	2	069635	1.28e+01

DR EMBL: Y08256; E283902; -
SQ SEQUENCE 526 AA; 56210 MW; 5AD5F9A0 CRC32;

Query Match 56.9%; Score 66; DB 1; Length 526;
Best Local Similarity 52.9%; Pred. No. 4,09e+02;
Matches 9; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Db 220 FGGFGAFHNAQTSSQ 236
1 FGGFTGARKSARKLANQ 17

RESULT 3
ID 060719 PRELIMINARY; PRT; 776 AA.
AC 060719:
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE CULLIN 1.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-CERVIX;
RA MICHEL J.J., XIONG Y.;
RL CELL GROWTH DIFFER. 0:0-0(1998).
DR EMBL: AF062536; G3139077;
DR PROSITE: PS01256; CULLIN_1;
SQ SEQUENCE 776 AA; 89678 MW; 480E807E CRC32;

Query Match 51.7%; Score 60; DB 4; Length 776;
Best Local Similarity 46.7%; Pred. No. 7,95e+01;
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;
Db 566 FTATYASRHSGRKLT 580
1 FGGFTGARKSARKLANQ 15

RESULT 4
ID 028463 PRELIMINARY; PRT; 177 AA.
AC 028463:
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 20.6 KD PROTEIN.
GN AF1812
OS ARCHAEOLOBUS FULGIDUS.
OC ARCHAEBACTERIA; EURYARCHAEOTA; ARCHAEOGLOBALES; ARCHAEOGLOBACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE: 98049343
RA KLENN H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,
RA KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,
RA RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KIRPIDES N.C.,
RA FLEISCHMANN R.D., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,
RA KIRKNESS E.F., REICH C.I., MCNEIL L.K., BADGER J.H., GLODER A., ZHOU L.,
RA OVERBEER R., GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTERBERCK T.,
RA COTTON M.D., SPRIGGS T., ARTIACH P., KAINE B.P., SYKES S.M.,
RA SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
RA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., MOESE C.R.,
RA VENTER J.C.;
RL NATURE 390:364-370(1997).
DR EMBL: AE000978; G2648748;
DR TIGR: AF1812;
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 177 AA; 20596 MW; 9721D71F CRC32;

Query Match 50.9%; Score 59; DB 1; Length 177;
Best Local Similarity 46.7%; Pred. No. 1,28e+00;

Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Db 138 FLVGSIRKARLTG 152
1 FGGFTGARKSARKLANQ 15

RESULT 5
ID 066484 PRELIMINARY; PRT; 211 AA.
AC 066484:
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE RIBOSOMAL PROTEIN S04.
GN RPSD.
OS ACUIPEX AEOLICUS.
OC EUBACTERIA; ACUIFICALES; ACUIFICACEAE; ACUIPEX.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VF5;
RA DECKERT G., WARREN P.V., GASTERLAND T., YOUNG W.G., LENOX A.L.,
RA GRAHAM D.E., OVERBEER R., SNEAD M.A., KELLER M., AUCAY M., HUBER R.,
RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
RL NATURE 392:353-358(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-VF5;
RA DECKERT G., WARREN P.V., GASTERLAND T., YOUNG W.G., LENOX A.L.,
RA GRAHAM D.E., OVERBEER R., SNEAD M.A., KELLER M., AUCAY M., HUBER R.,
RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AE000672; G2962819;
DR PROSITE: PS00632; RIBOSOMAL_S4; 1.
KW RIBOSOMAL PROTEIN.
SQ SEQUENCE 211 AA; 24815 MW; 8A762049 CRC32;

Query Match 50.9%; Score 59; DB 2; Length 211;
Best Local Similarity 40.0%; Pred. No. 1,28e+00;
Matches 6; Conservative 5; Mismatches 4; Indels 0; Gaps 0;
Db 108 GPASTRQARQOLVAH 122
3 GFTGARKSARKLANQ 17

RESULT 6
ID 099899 PRELIMINARY; PRT; 133 AA.
AC 099899:
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE P65-MEMBER OF THE STEROID/THYROID RECEPTOR SUPERFAMILY (FRAGMENT).
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE: 96253399.
RA HANAUER M., SZEMERAJ J., ADAMS A.K., WALASZEK Z.;
RL CANCER DETECT. PREV. 20:94-102(1996).
DR EMBL: S82307; G1839614;
DR PFMW: PF00105; zf-C4.
FT NON_TER 133
SQ SEQUENCE 133 AA; 15100 MW; B7F06493 CRC32;

Query Match 49.1%; Score 57; DB 4; Length 133;
Best Local Similarity 53.8%; Pred. No. 3,27e+00;
Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 28 GGRAGRRGARRL 40
2 GGTGARKSARKLANQ 14


```

RESULT 7
ID 041349 PRELIMINARY; PRT: 152 AA.
AC 041349;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE GLYCINE-RICH PROTEIN
OS LYCOPERSICON ESCULENTUM (TOMATO).
OC EUPHARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE;
CC SOLANALES; SOLANACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV VF36;
RA SMITH A.G.;
RL SUBMITTED (OCT-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; L47626; G1006565;
SQ SEQUENCE 152 AA; 15284 MW; 752A9F05 CRC32;

Query Match 49.1%; Score 57; DB 10; Length 152;
Best Local Similarity 58.3%; Pred. No. 3.27e+00;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Db 51 FGGFGVGRGAIR 62
OY 1 FGGFGARSAAR 12

RESULT 8
ID 043245 PRELIMINARY; PRT: 614 AA.
AC 043245;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DE PROTEIN.
OS HOMO SAPIENS (HUMAN).
OC EUPHARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
CC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BREAST;
RA HANAUSEK M.E.;
RL SUBMITTED (OCT-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-BREAST;
RA HANAUSEK M.E.; SZEMRAJ J.Z., ADAMS A.K., WALASZEK Z.A.;
RL CANCER DETECT. PREV. 20.94-102(1996).
DR EMBL; AJ002425; F1215340;
SQ SEQUENCE 614 AA; 68629 MW; 1EA2C7A9 CRC32;

Query Match 49.1%; Score 57; DB 4; Length 614;
Best Local Similarity 53.8%; Pred. No. 3.27e+00;
Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 292 GGRAGPRGARRL 304
OY 2 GGTGARKSARKL 14

RESULT 9
ID 013436 PRELIMINARY; PRT: 801 AA.
AC 013436;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DE NUCLEOLAR RNA HELICASE GU (FRAGMENT).
OS HOMO SAPIENS (HUMAN).
OC EUPHARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
CC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE: 9618853.
RA VALDEZ B.C., HENNING D., BUSCH R.K., WOODS K., FLORES-ROZAS H.,
RA HURWITZ J., PERIAKI L., BUSCH H.;
RL NUCLEIC ACIDS RES. 24:1220-1224(1996).
DR EMBL; U41387; G1230564;
DR PFAM; PF00270; DEAD.
DR PFAM; PF00271; helicase_C.
FT NON_TER 1
SQ SEQUENCE 801 AA; 89250 MW; F8F6B53C CRC32;

Query Match 49.1%; Score 57; DB 4; Length 801;
Best Local Similarity 41.2%; Pred. No. 3.27e+00;
Matches 7; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Db 739 YGFGRGGRGGRGQ 755
OY 1 FGGFGARSAKLANQ 17

RESULT 10
ID 004987 PRELIMINARY; PRT: 1061 AA.
AC 004987;
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DE ER-TYPE CALCIUM-TRANSPORTING ATPASE (EC 3.6.1.38).
GN ECAL OR ACA3.
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC EUPHARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE;
CC CAPPARALES; CRUCIFERAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RA LIANG F., SZE H.;
RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA HONG B., WANG Y., YOUNG J., SUSSMAN M.R., HARPER J.F.;
RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U96455; G2078292;
DR EMBL; U93845; G1943751;
DR PROSITE; PS00154; ATPASE_E1_E2; 1.
DR PFAM; PF00122; E1-E2_ATPASE.
KM HYDROLASE; TRANSMEMBRANE; PHOSPHORYLATION; ATP-BINDING.
FT MOD_RES 383
SQ SEQUENCE 1061 AA; 116365 MW; E5B8F256 CRC32;

Query Match 49.1%; Score 57; DB 10; Length 1061;
Best Local Similarity 47.1%; Pred. No. 3.27e+00;
Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Db 621 FVGFGVGRDPPRKEVRO 637
OY 1 FGGFGARSAKLANQ 17

RESULT 11
ID P93024 PRELIMINARY; PRT: 890 AA.
AC P93024;
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DE IAA24 (FRAGMENT).
GN IAA24.
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC EUPHARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE;
CC CAPPARALES; CRUCIFERAE.
RN [1]
RP SEQUENCE FROM N.A.
RA HARTER R.J.W., KIM J., THEOLOGIS A.;
RL PROC. NATL. ACAD. SCI. U.S.A. 94:11786-11791(1997).
DR EMBL; U79557; G2708484;

```



```

FT NON_TER 1 1
SO SEQUENCE 890 AA; 98323 MW; DC5A2F01 CRC32;
Query Match 48.3%; Score 56; DB 10; Length 890;
Best Local Similarity 53.8%; Pred. No. 5.18e+00;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 159 GGFSPRRRAEKL 171
|||:|:|:|
2 GGFGARSAKL 14

RESULT 12
ID 064965; PRELIMINARY; PRT; 902 AA.
AC 064965;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DE 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE TRANSCRIPTION FACTOR.
GN MF.
OS ABABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE;
OC CAPRARIACE; CRUCIFERAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA.
RA HARDTKE C.S., BERLETH T.;
RL EMBO J. 17:1405-1411(1998).
DR EMBL; AF037229; G2982222; -.
SQ SEQUENCE 902 AA; 99662 MW; 667EBA47 CRC32;

Query Match 48.3%; Score 56; DB 10; Length 902;
Best Local Similarity 53.8%; Pred. No. 5.18e+00;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 171 GGFSPRRRAEKL 183
|||:|:|:|
2 GGFGARSAKL 14

RESULT 13
ID 045760; PRELIMINARY; PRT; 1385 AA.
AC 045760;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE DELTA-ENDOTOXIN (FRAGMENT).
GN CRVA.
OS BACILLUS THURINGIENSIS.
OC PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PS17;
RA NARVA K.E., PAYNE J.M., SCHWAB G.E., HICKLE L.A., GALASAN T.,
RA STICK A.J.;
RL SUBMITTED (DEC-1992) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; L07025; G142869; -.
DR PFAM; PF00555; endotoxin.
FT NON_TER 1385 1385
SQ SEQUENCE 1385 AA; 152439 MW; 2592A8C7 CRC32;

Query Match 48.3%; Score 56; DB 2; Length 1385;
Best Local Similarity 50.0%; Pred. No. 5.18e+00;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Db 943 FGDEKALRLVNO 956
|:|:|:|
4 FTGARSAKL 17

RESULT 14
ID 040346; PRELIMINARY; PRT; 470 AA.
AC 040346;

```

```

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE LEGUMIN PRECUR PRECURSOR (FRAGMENT).
OS MAGNOLIA SALICIFOLIA
OC EUKARYOTA; PLANTA; EMBRYOIONTA; MAGNOLIOPHYTA; MAGNOLIOPSIDA;
OC MAGNOLIACE; MAGNOLIALES; MAGNOLIACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SINGLE TREE IN MUNICH BOTANICAL GARDEN; TISSUE-SEED;
RX MEDLINE; 95278215.
RA FISCHER H., HAAKE V., HORSTMANN C., JENSEN U.;
RL EUR. J. BIOCHEM. 229:645-650(1995).
DR EMBL; X82463; G793854; -.
DR PFAM; PF00190; Seedstore_11s.
KW SIGNAL.
FT NON_TER 1 1
FT SIGNAL <1 19 POTENTIAL.
SQ SEQUENCE 470 AA; 52986 MW; EEB11097 CRC32;

Query Match 47.4%; Score 55; DB 10; Length 470;
Best Local Similarity 46.7%; Pred. No. 8.16e+00;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 236 AFGVRETKRLQSO 250
|:|:|:|:|
3 GFTGARSAKL 17

RESULT 15
ID 040347; PRELIMINARY; PRT; 476 AA.
AC 040347;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE GLOBULIN PRECURSOR.
OS MAGNOLIA SALICIFOLIA.
OC EUKARYOTA; PLANTA; EMBRYOIONTA; MAGNOLIOPHYTA; MAGNOLIOPSIDA;
OC MAGNOLIACE; MAGNOLIALES; MAGNOLIACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SINGLE TREE IN MUNICH BOTANICAL GARDEN; TISSUE-SEED;
RX MEDLINE; 95278215.
RA FISCHER H., HAAKE V., HORSTMANN C., JENSEN U.;
RL EUR. J. BIOCHEM. 229:645-650(1995).
DR EMBL; X82464; G793856; -.
DR PFAM; PF00190; Seedstore_11s.
KW SIGNAL.
FT SIGNAL.
FT SIGNAL 1 23 POTENTIAL.
SQ SEQUENCE 476 AA; 53663 MW; 5435E3FC CRC32;

Query Match 47.4%; Score 55; DB 10; Length 476;
Best Local Similarity 46.7%; Pred. No. 8.16e+00;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 240 AFGVRETKRLQSO 254
|:|:|:|:|
3 GFTGARSAKL 17

Search completed: Fri Apr 16 14:19:42 1999
Job time : 32 secs.

```

 Release 3.1a John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

 (TM)

Release 3.1a John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Mparch_tpr n.a. n.a. Smith-Waterman search, using a protein query
 which has been backtranslated into n.a. using IUPAC symbols

Run on: Tue Apr 27 10:17:01 1999; Maspar time 25.49 Seconds

Tabular output not generated. 272.182 Million cell updates/sec

Title: >US-09-011-797-2
 Description: (1-17) from US09011797.pep
 Perfect Score: 200
 N.A. Sequence: 1 TTYGNGNGNTTYACNGNGC.....CNGNNAATYNGCAAYCAR 51
 Comp: AARCCNCAARATGNCNCG.....GNKNTTYRANCNTRNGTY

Scoring table: TABLE bkttranslated
 Gap 40

Nmatch STD : Dbase 0: Query 0

Searched: 188442 segs, 68026449 bases x 2

Post-processing: Minimum Match 08
 Listing first 45 summaries

Database:

n:geneg32
 1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
 8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
 14:part14 15:part15 16:part16 17:part17 18:part18
 19:part19 20:part20 21:part21 22:part22 23:part23
 24:part24 25:part25 26:part26 27:part27 28:part28
 29:part29 30:part30 31:part31 32:part32 33:part33
 34:part34 35:part35 36:part36 37:part37 38:part38
 39:part39 40:part40

Statistics: Mean 34.207; Variance 134.692; scale 0.254

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	196	98.0	932	34	T79888 DNA encoding human op	5.72e-08
2	106	53.0	1656	33	HMG-CoA reductase deg	5.00e+00
3	96	48.0	3411	40	DNA encoding a Bacill	3.07e+01
4	96	48.0	580073	27	Mycoplasma genitalium	3.67e+01
5	95	47.5	359	7	Active site polynucle	3.67e+01
6	91	45.5	297	8	Human brain Expressed	7.41e+01
7	91	45.5	1302	38	Magnetostrictum sp.	7.41e+01
8	91	45.5	1606	3	Mec-4 gene.	7.41e+01
9	91	45.5	1607	6	Mec-4 gene.	7.41e+01
10	91	45.5	3471	40	DNA encoding a Bacill	7.41e+01
11	91	45.5	36335	16	Ad2/-ORF6/PGK-CFR nu	7.41e+01
12	90	45.0	3872	17	Sockeye salmon growth	8.82e+01

RESULT	ID	Score	Query Match	Length	DB ID	Description	Pred. No.
1	T79888	98.0	932	34	T79888	DNA encoding human op	5.72e-08
2	T79888	106	53.0	1656	33	HMG-CoA reductase deg	5.00e+00
3	T79888	96	48.0	3411	40	DNA encoding a Bacill	3.07e+01
4	T79888	96	48.0	580073	27	Mycoplasma genitalium	3.67e+01
5	T79888	95	47.5	359	7	Active site polynucle	3.67e+01
6	T79888	91	45.5	297	8	Human brain Expressed	7.41e+01
7	T79888	91	45.5	1302	38	Magnetostrictum sp.	7.41e+01
8	T79888	91	45.5	1606	3	Mec-4 gene.	7.41e+01
9	T79888	91	45.5	1607	6	Mec-4 gene.	7.41e+01
10	T79888	91	45.5	3471	40	DNA encoding a Bacill	7.41e+01
11	T79888	91	45.5	36335	16	Ad2/-ORF6/PGK-CFR nu	7.41e+01
12	T79888	90	45.0	3872	17	Sockeye salmon growth	8.82e+01

Human myod gene and m
 elc cDNA.
 Pseudomonas aeruginos
 Total base sequence o
 Total base sequence o
 Kaposi's sarcoma asso
 Human gene signature
 Aspergillus nidulans
 Sequence encoding pol
 5' region of bacterial
 M. tuberculosis cellu
 Doc2 (brain-specific
 Genomic sequence enco
 Alpha-1B adrenergic r
 Alpha 1b adrenergic r
 MCF FKN provirus env
 Sequence of ACP (acv)
 Aureobasidin sensiti
 Human host cell prote
 Murine puromycin-sens
 Aspergillus niger cat
 catr gene.
 Human puromycin-sensi
 Human puromycin-sensi
 Human endocytelin-2 ge
 Human puromycin-sensi
 Aureobasidin resistan
 HIV-2 variant HIV-20
 Complete sequence of
 DNA encoding genes in
 Construct ECL2 (Conla
 Kaposi's sarcoma asso
 Kaposi's sarcoma asso


```

Db      Query March          98.0%   Score 196; DB 34; Length 932;
        Best Local Similarity 56.9%; Pred. No. 5,72e-08;
        Matches 29; Conservative 11; Mismatches 11; Indels 0; Gaps 0;

Oy      1    ttgggagcttcactggcgccgcgaagtcacgcccgaaagtgtgcacaaccg 414
         ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
         1 TTYGGNGNTTTCACGNGCNGMNAARMSNCGNMGNARYTNGCNAAVCAR 51

RESULT 2
ID      T85271 standard; DNA; 1656 BP.
AC      T85271.
DE      08-DEC-1997 (first entry)
DR      HMG-CoA reductase degradation polypeptide 1 HRD1 gene.
KW      3-hydroxy-3-methylglutaryl: coenzyme A; cholesterol; Hrd1p; Hrd2p;
OS      Hrd3p: hypercholesterolaemia; yeast; ss.
FH      Saccharomyces cerevisiae.
FT      Key Location/Qualifiers
       cds 1..1656
              /*tag= a
              /product= Hrd1p

FN      W09707219-A2.
PD      27-FEB-1997.
PE      16-AUG-1996; IB1161.
PR      17-AUG-1995; US-002381.
PA      (REGC ) UNIV CALIFORNIA.
PI      Hampton R. Rine JD;
DR      WPI. 97-165303/15.
OR      P-P5DB; W27150.
PT      3-Hydroxy-3-methyl:glutaryl CoA reductase degradation polypeptide(s)
PS      - useful as therapeutic agents to reduce hypercholesterolaemia
CC      Claim 2; Page 88-89; 132pp: English.
CC      The present sequence represents the 3-hydroxy-3-methylglutaryl
CC      (HMG)-CoA reductase degradation (HRD) protein HRD1 gene. The HRD1 gene
CC      encodes the Hrd1 protein. Hrd proteins can be used to regulate the
CC      degradation of HMG-CoA reductase, e.g. as therapeutic agents to reduce
CC      hypercholesterolaemia, and to elucidate how the cholesterol pathway
CC      modulates the degradation of HMG-CoA reductase. In addition, as a result,
CC      of their ability to bind the proteasome complex, antibodies that
CC      specifically bind hrd polypeptides can be used to isolate the proteasome
CC      complex. Further, they can be used in various assays to identify
CC      compounds that modify the degradation of HMG-CoA reductase independently
CC      of the beneficial IDL receptor control axis. The nucleic acid molecules
CC      can be used as molecular probes for the isolation of homologous nucleic
CC      acid molecules and for the detection of HRD nucleic acid molecules in
CC      yeast.
SQ      Sequence 1656 BP; 537 A; 324 C; 304 G; 491 T;

Query Match          53.0%; Score 106; DB 33; Length 1656;
Best Local Similarity 46.2%; Pred. No. 5.00e+00;
Matches 18; Conservative 8; Mismatches 13; Indels 0; Gaps 0;

Db      1555 acggggcgcgagcaaatgccagaanaattgtcataccag 1633
         ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
         Oy 13 ACNGNGCNGMNAARMSNCGNMGNARYTNGCNAAVCAR 51

RESULT 3
ID      V16518 standard; DNA; 3411 BP.
AC      V16518.
AD      V16518.
DE      11-JUN-1998 (first entry)
DR      DNA encoding a Bacillus thuringiensis toxin.
KW      Toxin; lepidopteran pest; control; Agrotis ipsilon; black cutworm;
OS      Heliothis virescens; Helicoverpa zea; ss.
FH      Bacillus thuringiensis.
FT      Key Location/Qualifiers
       cds 1..3405
              /*tag= a

FN      W09800546-A2.
PD      08-JAN-1998.
PE      01-JUL-1987; U11658.
PR      01-JUL-1996; US-674002.

```

PA	(MVCO)MYCOGEN CORP.
PI	Narva KE, Schnepf HE, Stockhoff BA, Walz M, Wlcker C;
DR	MPL: 98-086971/08.
PT	P-FSDB: W46859.
FR	New isolated Bacillus thuringiensis isolate(s) - used to obtain
PR	genes encoding toxins which are active against lepidopteran pests
PS	such as the Black cutworm
CC	Example 8; Pages 106-108; 183pp; English.
CC	The present sequence is isolated from a Bacillus thuringiensis isolate
CC	(HJ129). The gene product was expressed as a fusion protein with
CC	the first 28 codons of GYIAC in a Pseudomonas expression system. The
CC	spectification describes B. thuringiensis toxins which are active
CC	against Lepidoptera pests. The toxin isolates can be used for the
CC	control of lepidopteran pests such as Agrotis ipsilon (black cutworm),
CC	Heliothis virescens and Helicoverpa zea. PCR primers and probes can be
CC	derived from the polynucleotides encoding the toxins and used for the
CC	amplification and detection of other toxin-encoding sequences.
SQ	Sequence 3411 BP; 1134 A; 365 C; 740 G; 952 T;
Df	Query Match 48.0%; Score 96; DB 40; Length 3411;
Bt	Best Local Similarity 40.5%; Pred. No. 3.07e+01;
Matches	15; Conservative 9; Mismatches 13; Indels 0; Gaps 0
Oy	2231 cacttcagctacgtcacgtagccaggaaattatcccac 2267 :: :: : : :: : :: : 12 YACNGGCGNMGNARMSNCGCMGARVTCNCNPAY 48
ID	T58840 standard; DNA; 580073 BP.
DE	Mycoplasma genitalium genome.
DI	27-MAR-1997 (first entry)
DM	KW Megabase Shotgun sequencing method; open reading frame; ORF; ss. OS Mycoplasma genitalium.
FH	Key location/Qualifiers
FT	cds 8552..9184
FT	/tag= a
FT	/label= MG006
FT	/note= "Previously identified as MORF-20076, the
FT	encoded protein shows 27.59 percentage
FT	identity to thymidylate kinase (CDC8)
FT	from Saccharomyces cerevisiae"
FT	11252..12040
FT	cds
FT	/tag= b
FT	/label= MG009
FT	/note= "Previously identified as MORF-20078, the
FT	encoded protein shows 35.43 percentage
FT	identity to the Bacillus subtilis hypothetical
FT	protein covered in accession number
FT	GB:D26185_102
FT	12069..12725
FT	cds
FT	/tag= c
FT	/label= MG010
FT	/note= "Previously identified as MORF-20079, the
FT	encoded protein shows 25.73 percentage
FT	identity to DNA primase (dnaE) from
FT	Clostridium acetobutylicum"
FT	complement (13570..14247)
FT	/tag= d
FT	/label= MG012
FT	/note= "Previously identified as MORF-20080, the
FT	encoded protein shows 31.50 percentage
FT	identity to the ribosomal protein S6
FT	modification protein (rimK) from Escherichia
FT	coli"
FT	complement (14396..15217)
FT	/tag= e
FT	/label= MG013
FT	/note= "Previously identified as MORF-19823, MORF-20080
FT	and MORF-20081, the encoded protein shows 33.04
FT	percentage identity to 5,10-methylene-tetra-


```
FT      cds      hydrofolate dehydrogenase (fold) from E. coli"
FT      17474..19243
FT      /*tag- f
FT      /label- MG015
FT      /note- "Previously identified as MORF-20084, the
FT      encoded protein shows 32.23 percentage
FT      identity to transport ATP-binding protein
FT      (meba) from E. coli"
FT      26478..27344
FT      cds      /*tag- g
FT      /label- MG023
FT      /note- "Previously identified as MORF-20092, the
FT      encoded protein shows 45.96 percentage
FT      identity to fructose-bisphosphate aldolase
FT      (far) from B. subtilis"
FT      27345..28448
FT      cds      /*tag- h
FT      /label- MG024
FT      /note- "Previously identified as MORF-19826 and
FT      MORF-20093, the encoded protein shows 46.84
FT      percentage identity to GTP-binding protein
FT      from E. coli"
FT      36987..38978
FT      cds      /*tag- i
FT      /label- MG032
FT      /note- "Previously identified as MORF-20099, the
FT      encoded protein shows 26.82 percentage
FT      identity to ATP-dependent nuclease (adda)
FT      from B. subtilis"
FT      39242..39904
FT      cds      /*tag- j
FT      /label- MG033
FT      /note- "Previously identified as MORF-20100, the
FT      encoded protein shows 35.90 percentage
FT      identity to glycerol uptake facilitator
FT      (gipf) from B. subtilis"
FT      complement (39873..40514)
FT      cds      /*tag- k
FT      /label- MG034
FT      /note- "Previously identified as MORF-20101, the
FT      encoded protein shows 48.13 percentage
FT      identity to thymidylate kinase (cdk)
FT      from B. subtilis"
FT      40543..41787
FT      cds      /*tag- l
FT      /label- MG035
FT      /note- "Previously identified as MORF-20102, the
FT      encoded protein shows 30.71 percentage
FT      identity to histidyl-tRNA synthetase (hiss)
FT      from Mycobacterium leprae"
FT      complement (44751..46277)
FT      cds      /*tag- m
FT      /label- MG038
FT      /note- "Previously identified as MORF-20105, the
FT      encoded protein shows 46.83 percentage
FT      identity to glycerol kinase (gipk)
FT      from E. coli"
FT      complement (46268..47422)
FT      cds      /*tag- n
FT      /label- MG039
FT      /note- "Previously identified as MORF-19831 and
FT      MORF-20106, the encoded protein shows 43.20
FT      percentage identity to glycerol-3-phosphate
FT      dehydrogenase (gdt2) from S. cerevisiae"
FT      49377..49643
FT      cds      /*tag- o
FT      /label- MG041
FT      /note- "The encoded protein shows 48.86 percentage
FT      identity to phosphohistidinoprotein-hexose
FT      phosphotransferase (psh) from Mycoplasma
FT      capricolium"
FT      50060..51520
FT      /*tag- p

FT      /label- MG042
FT      /note- "Previously identified as MORF-19832 and
FT      MORF-20108, the encoded protein shows 41.92
FT      percentage identity to spermidine/
FT      putrescine transport ATP-binding protein
FT      (potA) from E. coli"
FT      51525..52382
FT      cds      /*tag- q
FT      /label- MG043
FT      /note- "Previously identified as MORF-20110, the
FT      encoded protein shows 26.51 percentage
FT      identity to spermidine/putrescine transport
FT      system permease protein (pots) from E. coli"
FT      52366..53220
FT      cds      /*tag- r
FT      /label- MG044
FT      /note- "Previously identified as MORF-20111, the
FT      encoded protein shows 29.45 percentage
FT      identity to spermidine/putrescine transport
FT      system permease protein c (potc) from E. coli"
FT      54658..55605
FT      cds      /*tag- s
FT      /label- MG046
FT      /note- "Previously identified as MORF-20112, the
FT      encoded protein shows 36.60 percentage
FT      identity to sialoglycoprotease (gcp)
FT      from Pasteurella haemolytica"
FT      complement (56970..58310)
FT      cds      /*tag- t
FT      /label- MG048
FT      /note- "Previously identified as MORF-19834,
FT      MORF-20114 and MORF-20115, the encoded protein
FT      shows 43.02 percentage identity to signal
FT      recognition particle protein (ifh) from B.
FT      subtilis"
FT      58117..59079
FT      cds      /*tag- u
FT      /label- MG049
FT      /note- "Previously identified as MORF-20114 and
FT      MORF-20115, the encoded protein shows 44.78
FT      percentage identity to purine-nucleoside
FT      phosphorylase (deod) from E. coli"
FT      59083..59754
FT      cds      /*tag- v
FT      /label- MG050
FT      /note- "Previously identified as MORF-20117, the
FT      encoded protein shows 83.03 percentage
FT      identity to deoxyribose-phosphate aldolase
FT      (deoc) from Mycoplasma pneumoniae"
FT      complement (64898..65731)
FT      cds      /*tag- w
FT      /label- MG056
FT      /note- "Previously identified as MORF-20122, the
FT      encoded protein shows 30.25 percent
FT      identity to the protein disclosed in
FT      GB:D26185..99 from B. subtilis"
FT      complement (65713..66249)
FT      cds      /*tag- x
FT      /label- MG057
FT      /note- "Previously identified as MORF-20123, the
FT      encoded protein shows 38.90 percent
FT      identity to the protein disclosed in
FT      GB:D26185..104 from B. subtilis"
FT      81047..82597
FT      cds      /*tag- y
FT      /label- MG067
FT      /note- "Previously identified as MORF-19845, the
FT      encoded protein shows 28.84 percentage
FT      identity to glutamic acid specific protease
FT      (spase) from Staphylococcus aureus"
FT      91065..91919
FT      /*tag- z
FT      /label- MG070
```


FT /note= "Previously identified as MORF-20136, the
 FT encoded protein shows 34.8 percentage
 FT identity to ribosomal protein S2 (rps2)
 FT from *Spirulina plantensis*"
 FT 103104.104324
 FT /tag= aa
 FT /label= MG077
 FT /note= "Previously identified as MORF-20140, the
 FT encoded protein shows 28.05 percentage
 FT identity to oligopeptide transport system

... Note: remainder of annotations omitted.

Query Match 48.0%; Score 96; DB 27; Length 580073;
 Best Local Similarity 43.2%; Pred. No. 3.07e+01;
 Matches 19; Conservative 8; Mismatches 17; Indels 0; Gaps 0;
 Db 387537 tggttaccattgtctgacctcttctgtgtgcaaac 387580
 Cp 50 TGRITNGCNARYTTNCKNGCNSWYTTNCKNGCNCNGTRANCC 7

RESULT 5
 ID 046579 standard; DNA; 359 BP.
 AC 046579;
 DT 22-DEC-1993 (first entry)
 DE Active site polynucleotide fragment.
 KM Active site; active region; prediction; secondary structure;
 OS energy value; engineering; ss.
 PN Synthetic.
 PS J05155899-A.
 PD 22-JUN-1993.
 PE 23-AUG-1991; 212284.
 PR 23-AUG-1991; JP-212284.
 PA (SAGA) SAGA1 CHEM RES CENTRE.
 DR WPI; 93-232353/29.
 PT Prediction of the active site in physiologically active
 PT polypeptide - by determ. of sec. structure energy values of
 PT partial regions along prim. sequence of physiologically active
 PT polynucleotide, etc.
 PS disclosure: Page 23-24; 43pp; Japanese.
 CC This sequence represents a polynucleotide which was used to
 CC demonstrate the method of the invention. This oligomer represents
 CC an active site/region and the method of the invention may be used to
 CC predict the active site of a polynucleotide by determining the energy
 CC values of the secondary structure of various partial regions of the
 CC polynucleotide to be tested and the primary sequence as well. At
 CC least one partial region should show a local peak energy value
 CC significantly higher or lower than the average standard energy value.
 CC This method may be used in the elucidation of the mechanism of
 CC action of polynucleotides or polypeptides, and in the engineered
 CC improvement of such sequences.
 SQ Sequence 359 BP; 69 A; 87 C; 110 G; 93 T;
 Query Match 47.5%; Score 95; DB 7; Length 359;
 Best Local Similarity 44.4%; Pred. No. 3.67e+01;
 Matches 16; Conservative 7; Mismatches 13; Indels 0; Gaps 0;
 Db 52 caggtgcccgaagccaccgtaactagtgaacc 87
 Cp 14 CNGGNCNCMGNAARWNGCNGNARITNGCNAAYC 49

RESULT 6
 ID 060810 standard; DNA; 297 BP.
 AC 060810;
 DT 16-MAR-1994 (first entry)
 DE Human brain Expressed Sequence Tag EST00914.
 KM Gene transcription product; genetic markers; tagging; in vivo;
 OS Homo sapiens.
 PN M09316178-A.
 PD 19-AUG-1993.

PF 12-FEB-1993; U01294.
 PR 12-FEB-1993; US-837195.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
 PI Adams MD, Moreno RF, Venter CJ;
 DR WPI; 93-272882/34.
 PT Enriched oligonucleotides and corresp. sequences - used as
 PT markers for human genes transcribed in-vivo, facilitate tagging
 PT of most human genes
 PS Example 4; Page 394; 500pp; English.
 CC The Expressed Sequence Tag was isolated from a human brain cDNA
 CC library as part of a large set of ESTs which can be used as markers
 CC for human genes transcribed in vivo. They can be used to facilitate
 CC tagging of most human genes, for mapping locations of expressed genes
 CC on chromosomes, for individual or forensic identification, for mapping
 CC locations of disease-associated genes, for identification of tissue
 CC type, and for prep. of antisense sequences, probes and constructs.
 CC EST00914 has a "poor" coding probability as evaluated using the
 CC coding-region prediction program CRM. See also Q59041-Q61440.
 SQ Sequence 297 BP; 126 A; 44 C; 45 G; 82 T;

Query Match 45.5%; Score 91; DB 8; Length 297;
 Best Local Similarity 43.6%; Pred. No. 7.41e+01;
 Matches 17; Conservative 7; Mismatches 15; Indels 0; Gaps 0;
 Db 170 gattcactgtgtaagtaaatgtacacacaccttcaa 208
 Cp 8 GNTTYACNGNCNGMNAARWNGCNGNARITNGCNA 46

RESULT 7
 ID T90224 standard; DNA; 1302 BP.
 AC T90224;
 DT 08-MAY-1998 (first entry)
 DE Magnetospirillum sp. Maga DNA degenerate homologue.
 KM Maga; immobilisation; magnetic particle; magnetic bacteria; ss.
 OS Magnetospirillum sp.
 PN W09735964-A1.
 PD 02-OCT-1997.
 PE 27-MAR-1997; J01043.
 PR 16-MAY-1996; JP-146833.
 PR 27-MAR-1996; JP-097536.
 PA (DENK) TDK CORP.
 PI Kamiya S, Matsunaga T, Namba K;
 DR WPI; 97-489628/45.
 PT Proteins bound to magnetic particles - useful as immobilised enzymes
 PT and analytical reagents
 PS disclosure: Page 47-48; 70pp; Japanese.
 CC This sequence encodes a novel degenerate homologue of the Maga protein,
 CC which is used in a method of producing magnetic particles for the
 CC immobilisation of various substances. These particles may be used as
 CC substrates for immobilising enzymes, DNA carriers, a protein biosynthesis
 CC system, analytical reagents or to assay antigens or other substances. The
 CC amount of a substance in a sample can be measured using the products. The
 CC immobilised enzymes are easily prepared by culturing transformed magnetic
 CC bacteria, and isolating the magnetic particles.
 SQ Sequence 1302 BP; 143 A; 118 C; 259 G; 216 T;
 Query Match 45.5%; Score 91; DB 38; Length 1302;
 Best Local Similarity 26.5%; Pred. No. 7.41e+01;
 Matches 13; Conservative 11; Mismatches 25; Indels 0; Gaps 0;
 Db 999 rathgngartlywsntlytngngaracngngaargcngtaar 1047
 Cp 49 GRTTNGCNARYTTNCKNGCNSWYTTNCKNGCNCNGTRANCCGCAA 1

RESULT 8
 ID Q20264 standard; CDNA; 1606 BP.
 AC Q20264;
 DT 25-MAR-1992 (first entry)
 DE Mec-4 gene.
 KM Neurodegeneration; cell death; nematode; Huntington's; Lou Gehrig's;
 KW Alzheimer's; roundworm; ss.

OS	Caenorhabditis elegans.	Location/Qualifiers
FT	Key	14..1495
FT	cds	/*tag= a
FT	polya_site	1567..1606
FT		/*tag= b
PN	W0919007-A.	
PD	12-DEC-1991.	
PF	30-MAY-1991: U03826.	
PR	30-MAY-1990: US-530968.	
PA	(UYCO-) COLUMBIA UNIV NEW Y.	
P1	Chalfie M, Wolinsky E, Driscoll M;	
DR	WPI: 92-007488/01.	
DR	P-PSDB: R20107.	
PT	DNA sequences involved in neuronal degeneration - used for	
P1	detecting, diagnosing, treating and preventing degenerative	
PT	diseases.	
PS	Claim 10: Fig 14: 57pp: English.	
CC	The mec-4 (mechanosensory abnormal) gene encodes a protein associa-	
CC	ted with the late-onset degeneration of a specific set of six touch	
CC	receptor neurons. Strains of C. elegans having mutations in the	
CC	mec-4 gene (el611, u214, u231) have been identified in a screen for	
CC	touch sensitivity. No description of the cloning method is provid-	
CC	ed. The DNA may be useful for detecting, diagnosing, preventing and	
CC	treating degenerative diseases such as Alzheimer's disease, amy-	
CC	otrophic lateral sclerosis and Huntington's disease.	
CC	See also Q20763.	
SO	Sequence 1606 BP: 500 A: 312 C: 360 G: 434 T:	
Query Match	45.5%: Score 91: DB 3: Length 1606;	
Best Local Similarity	43.3%: Pred. No. 7.41e+01;	
Matches	13: Conservative 8: Mismatches 9; Indels 0; Gaps 0;	
Db	1308 gattgtcaactctctagcagatttggtg 1337	
Cp	49 GRTTGCNARTTNTCKNCNSMYTTCKNG 20	
RESULT	9	
ID	Q38934 standard: cDNA: 1607 BP.	
AC	Q38934;	
DT	20-JUL-1993 (first entry)	
DE	mec-4 gene.	
KW	deg-1; mec-4; nerve cells; neuronal cells; neurodegenerative;	
OS	disease; drug screening; testing; animal model.	
KM	Caenorhabditis elegans.	
FT	Key	Location/Qualifiers
FT	cds	15..1498
FT		/*tag= a
PN	US5196333-A.	
PD	23-MAR-1993.	
PF	30-MAY-1990: 530968.	
PR	30-MAY-1990: US-530968.	
PA	(UYCO) UNIV COLUMBIA NEW YORK.	
P1	Chalfie M, Driscoll M, Wolinsky E;	
DR	WPI: 93-116796/14.	
DR	P-PSDB: R33890.	
PT	Isolated nucleic acid molecule encoding wild type free living	
PT	nematode protein - comprises C. elegans deg-1 gene with specific	
PT	DNA sequence and mutated for neuronal cell degeneration	
PS	Claim 5: Fig 9: 35pp: English.	
CC	This sequence represents the mec-4 gene from C. elegans. Mutations	
CC	in the deg-1 (Q38933) and mec-4 genes cause mature functioning nerve	
CC	cells in C. elegans to die. This can provide an animal model to	
CC	study the causes of neurodegenerative diseases, and to cause the	
CC	degeneration of a diseased human cell. The C. elegans strains can	
CC	also be used to screen drugs to identify those which prevent or	
CC	decrease neuronal degeneration.	
SO	Sequence 1607 BP: 501 A: 312 C: 360 G: 434 T:	
Query Match	45.5%: Score 91: DB 6: Length 1607;	
Best Local Similarity	43.3%: Pred. No. 7.41e+01;	
Matches	13: Conservative 8: Mismatches 9; Indels 0; Gaps 0;	

```

Db 1308 gattgtaactgtcagccacttctgtg 1337
      ||| | ||| | | | | | | | | |
Cc 49 GRTTGCNARATTTCNKGNCNSWTTTCKNG 20

RESULT 10
ID ID V16515 standard; DNA; 3471 BP.
AC V16515;
DE 11-JUN-1998 (first entry)
KW DNA encoding a Bacillus thuringiensis toxin designated 86Bb1(a).
KW Toxin; lepidopteran pest; control; Agrotis ipsilon; black cutworm;
KW Heliothis virescens; Helicoverpa zea; ss.
OS Bacillus thuringiensis.
FH Key
FT CDS location/Qualifiers
      1..3471
      /*tag= a

FN MO9800546-A2.
FT 08-JAN-1998.
PE 01-JUL-1997; U11658.
PR 01-JUL-1996; US-674002.
PA (MYCO ) MYCOGEN CORP.
PI Marva KE, Schmepl HE, Stockhoff BA, Walz M, Wicker C;
DR WPI: 98-086971/08.
P-PSB: W46856.
PT New isolated Bacillus thuringiensis isolate(s) - used to obtain
PT genes encoding toxins which are active against lepidopteran pests
PT such as the Black cutworm
PS Example 5; Pages 86-88; 183pp; English.
CC The present sequence is isolated from a Bacillus thuringiensis isolate
CC (P866Bb1). It encodes a toxin designated 86Bb1(a) which is active
CC against lepidopteran pests. The toxin isolates can be used for the
CC control of lepidopteran pests such as Agrotis ipsilon (black cutworm),
CC Heliothis virescens and Helicoverpa zea. PCR primers and probes can be
CC derived from the polynucleotide encoding the toxin and used for the
CC amplification and detection of other toxin-encoding sequences.
CC Sequence 3471 BP; 1144 A; 591 C; 734 G; 1002 T;

Query Match 45.5%; Score 91; DB 40; Length 3471;
Best Local Similarity 43.3%; Pred. No. 7.41e+01;
Matches 13; Conservative 8; Mismatches 9; Indels 0; Gaps 0

Db 2304 gctagcaagtgcagagaattaccacac 2333
      || | ||| : : | | | : | | |
Cc 19 GCNMGNARMSNGCMGNARATTGNCNAY 48

RESULT 11
ID ID Q68003 standard; DNA; 36335 BP.
AC Q68003;
DE 26-MAR-1996 (first entry)
DE Ad2-/ORF6/PK-CRTR nucleotide sequence.
KW Recombinant adenovirus; Ad2/CRTR-1; adenovirus 2 serotype; Ela; E1b;
KW viral replication; gene expression; gene therapy; cystic fibrosis;
KW cystic fibrosis transmembrane conductance regulator; CRTR;
KW promoter; E3; p19; MHC; class I; viral latency; pulmonary airway; ds.
OS Synthetic.
FH Key
FT misc-feature location/Qualifiers
      12915..36335
      /*tag= a
      /note= "Represents residues 10676-34096 of Ad2-E4/ORF6"
FT misc-feature
      35069..35973
      /*tag= b
      /note= "Represents residues 33178-34082 of Ad2"
FT misc-feature
      12915..35054
      /*tag= c
      /note= "Represents residues 1-32815of Ad2"
FT exon
      28478..28790
      /*tag= d
      /number= 1
      /note= "33K protein"
FT intron
      28791..28992
      /*tag= e

```



```

FT      /number-1
FT      28993..29366
FT      /tag- f
FT      /number-2
FT      /note="33K protein"
FT      13279..14526
FT      /tag- g
FT      /product= 52,55K protein
FT      14547..16304
FT      /tag- h
FT      /product= iIra protein
FT      16331..16336
FT      /tag- i
FT      /note="Major late mRNA L1 poly A signal (putative)"
FT      16390..18105
FT      /tag- j
FT      /product= Penton protein
FT      18112..18708
FT      /tag- k
FT      /product= Pro-VII protein
FT      18778..19887
FT      /note="Precursor to major core protein"
FT      /tag- l
FT      /product= PV protein
FT      20188..20193
FT      /note="minor core protein"
FT      /tag- m
FT      /note="major late mRNA L2 poly-A signal (putative)"
FT      20240..20992
FT      /tag- n
FT      /product= pVI protein
FT      /note="Hexon associated precursor"
FT      21077..23983
FT      /tag- o
FT      /product= Hexon protein
FT      /note="Viron component II"
FT      24657..24662
FT      /tag- p
FT      /note="Major late mRNA L3 poly-A signal (putative)"
FT      complement (244729..26318)
FT      /tag- q
FT      /product= DBP protein
FT      /note="DNA binding or 72K protein"
FT      26347..28764
FT      /tag- r
FT      /product= 100K protein
FT      /note="Hexon assembled"
FT      29454..30137
FT      /tag- s
FT      /product= pVIII protein
FT      /note="Hexon-associated precursor"
FT      30444..30449
FT      /tag- t
FT      /note="major late mRNA L4 poly-A signal (putative)"
FT      31051..31530
FT      /tag- u
FT      /product= E3 19K protein
FT      /note="Glycosylated membrane protein"
FT      31707..32012
FT      /tag- v
FT      /product= E3 11.6K protein
FT      32008..32013
FT      /tag- w
FT      /note="E3-1 mRNA poly-A signal (putative)"
FT      33081..33086
FT      /tag- x
FT      /note="E3-2 mRNA poly-A signal (putative)"
FT      35013..35018
FT      /tag- y
FT      /note="major late mRNA L5 poly-A signal (putative)"
FT      1..12914
FT      misc-feature

```

```

FT      /tag- z
FT      /note="Represents residues 1-12914 of pad2/PGK-CFTR"
FT      380..914
FT      /tag- aa
FT      /note="pgk promoter"
FT      1011..5453
FT      /tag- ab
FT      /product= CFTR
FT      W09412649-A.
FT      PD 09-JUN-1994.
FT      PF 02-DEC-1993; U11667.
FT      PR 03-DEC-1992; US-985478.
FT      PR 01-OCT-1993; US-130682.
FT      PR 13-OCT-1993; US-136742.
FT      PA (GENZ) GENZYME CORP.
FT      PI Armentano D, Couture LA, Gregory RJ, Smith AE;
FT      DR WPT: 94-20027/24.
FT      PT Adeno:Virus-based gene therapy vectors - esp. useful for gene
FT      PT therapy of cystic fibrosis
FT      PS Example 15, Page 84-95; 167pp; English.
FT      CC This sequence represents the nucleotide sequence of the second generation
FT      CC adenovirus vector named Ad2-ORF6/PGK-CFTR. This virus lacks E1 and in its
FT      CC phase contains a modified transcription unit with the phosphoglycerate
FT      CC kinase (PGK) promoter and a poly A addition site flanking the cystic
FT      CC fibrosis transmembrane conductance regulator (CFTR) cDNA. The PGK
FT      CC promoter is only if moderate strength but it is long lasting and is not
FT      CC subject to shut off. The E4 region of the vector has also been modified
FT      CC in that the whole sequence has been removed and replace by ORF6, the
FT      CC only E4 gene essential for growth of adenovirus (Ad) in tissue culture.
FT      CC The DNA construct comprises a full length copy of the Ad2 genome from
FT      CC which the early region 1 genes (E1 genes) have been deleted and replaced
FT      CC by an expression cassette encoding CFTR. The expression cassette
FT      CC includes the promoter for PGK and a poly-A addition signal from the
FT      CC bovine growth hormone gene. The Ad2-ORF6/PGK-CFTR construct differs
FT      CC from that given in Q68802 (Ad2/CFTR-1), in that the latter utilises the
FT      CC endogenous E1a promoter, has no poly-A addition signal directly
FT      CC downstream of CFTR and retains an intact E4 region. This adenovirus may
FT      CC be administered to the pulmonary airways in the gene therapy of cystic
FT      CC fibrosis.
FT      SQ Sequence 36335 BP; 8597 A; 10000 C; 9786 G; 7952 T;
FT
FT      Query Match 45.5%; Score 91; DB 16; Length 36335;
FT      Best Local Similarity 42.5%; Pred. No. 7.41e+01;
FT      Matches 17; Conservative 8; Mismatches 15; Indels 0; Gaps 0;
FT
FT      Db 34861 acttacattacattgacatgacatgacatgacacaga 34900
FT      Cp 40 RYTNCKKNGCNSWYTTNCKNGCNCNGTIRANCCNCCRAA 1
FT
FT      RESULT 12
FT      ID T10156 standard; DNA; 3872 BP.
FT      AC T10156;
FT      DT 30-MAR-1996 (first entry)
FT      DE Sockeye salmon growth hormone type 2 gene OnGH2.
FT      KW Transgenic fish; salmonid; sockeye salmon; coho salmon;
FT      KW Atlantic salmon; vector; growth hormone type 1; somatotropin;
FT      KW ss.
FT      OS Oncorhynchus nerka.
FT      FH Key Location/Qualifiers
FT      FT promoter 1..288
FT      FT tata_signal 260..264
FT      FT exon 289..363
FT      FT /tag- c
FT      FT /label= Exon-1
FT      FT /codon_start= 354..356
FT      FT /tag- d
FT      FT /label= Intron-A
FT      FT exon 834..973
FT      FT /tag- e

```



```

FT      /tag= b
FT      /note= "3' end of promoter region"
FT      cds
FT      /tag= c
PN      MO9321347-A.
PD      28-OCT-1993.
PF      24-MAR-1993: U02767.
PR      10-APR-1992: US-866386.
PA      (FOXc-) FOX CHASE CANCER CENT.
PI      Emerson CP, Goldhamer DJ;
DR      WPI; 93-351756/44.
DR      P-PSDB: R42359.
PT      Transcription control element with enhancer activity - increases
PT      gene expression in myoblast(s) and is derived from upstream
PT      region of myod gene
PS      Claim 1; Page 16-18; 56pp; English.
CC      The sequence is that of the human myod gene and its upstream
CC      regulatory sequence. The transcriptional control element has
CC      enhancer activity in myogenic cells, causing increased expression of
CC      the myod gene. Antisense oligonucleotide fragments of the control
CC      sequence can be used for identifying and locating partic. functional
CC      regions in the control element. The myoblast-specific control
CC      element will aid advances in gene therapy using myoblast transfer
CC      and microinjection techniques. The control element may also be
CC      useful in biochemical assays for the activity of early transcription
CC      factors involved in regulating enhancers.
SQ      Sequence 4086 BP; 769 A; 1306 C; 1206 G; 805 T;

Query Match 45.0%; Score 90; DB 9; Length 4086;
Best Local Similarity 41.2%; Pred. No. 8,82e+01;
Matches 21; Conservative 8; Mismatches 22; Indels 0; Gaps

Db 901 ctgcgtcgccagctcgcggccctcgccacccacccggagaccctccca 951
Cp 51 YTGKTTGCGNARTTTCCKNCGNSWTTTCKNCGNCNGTANCCGCCRAA 1

RESULT 14
ID 053471 standard; cDNA; 4281 BP.
AC 053471:
DT 16-JUN-1994 (first entry)
DE ElK cDNA.
KW Lambda g11; expression vector; lambda-B1-ElK; protein tyrosine kinase
KW ElK; B1; Eph; subfamily; receptor-like tyrosine kinase; eph; eck;
KW phosphorylation; phosphotyrosine kinase insert domain; growth factor;
KW receptor kinase; platelet-derived growth factor receptor; SS.
OS Rattus rattus.
FH Key
FH Key Location/Qualifiers
FT cds 367..3321
FT FT /tag= a
FT FT /product= ElK

CA2083521-A.
PD 01-OCT-1993.
PF 23-NOV-1992: 083521.
PR 31-MAR-1992: US-861390.
PA (MOON ) MOUNT SINAI HOSPITAL CORP.
PI Letwin K, Pawson A, Reedijk M;
PI WPI; 93-406300/51.
DR P-PSDB: R44513.
PT Expression of phosphotyrosylated exogenous protein - in host cells
PT transformed with two vectors, one for the protein, the other for
PT catalytic domain of protein kinase
PS Disclosure; Fig 3: 55pp; English.
CC This sequence represents the elK cDNA which encodes the protein
CC tyrosine kinase, ElK. The ElK gene, B1, encode a protein which is
CC a member of the Eph subfamily of protein tyrosine kinases. The ElK
CC product is very similar to two other receptor-like tyrosine kinases,
CC eph and eck. Lambda-B1-ElK may be used in the production of
CC phosphotyrosylated exogenous protein along with a further vector encoding
CC the desired exogenous protein. These plasmid may be used to produce
CC phosphorylated proteins in host cells which have no intrinsic capacity
CC for phosphorylation, eg. bacteria. The system may be used for the

```


ACCESSION	AA867614	sequence.
NID	Q2963059	EST.
KEYWORDS		house mouse.
SOURCE		Mus musculus.
ORGANISM		Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
REFERENCE		Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS		1 (bases 1 to 488)
		Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
		Gaisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
		Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
		Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
		Waterston,R.
TITLE		The WashU-HMI Mouse EST Project
JOURNAL		Unpublished (1996)
COMMENT		Contact: Marra/Mouse Est Project WashU-HMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@wustl.wustl.edu This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Seq primer: -28ml3 rev2 ET from Amersham High quality sequence stop: 484. Location/Qualifiers 1..488 /organism="Mus musculus" /strain="C57Bl/6J" /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dt) primer [5' TGTTACCATCTGTGAGCGAGCGCCGCCGTTTTTTTTTTTTTTTTTT 3']; double stranded cDNA was ligated to Eco RI adaptors (pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. RNA provided by Dr. Bertrand Jordan. Library went through two rounds of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo." /db_xref="taxon:10090" /clone_1ib="1264210" /clone_1lb="Soares 2Nbxr" /sex="male" /tissue_type="Thymus" /dev_stage="4 weeks" /lab_host="DH10B" BASE COUNT 110 a 95 c 117 g 164 t 2 others ORIGIN
Query Match	58.0%;	Score 116; DB 13; Length 488;
Best Local Similarity	48.5%;	Pred. No. 1.74e-05;
Matches	16; Conservative	9; Mismatches 8; Indels 0; Gaps 0;
Db	45 TTGGTTTCCTGCTTGCCTGCTCCTTCTGCTACG 77	
Cp	51 YTGRTTNGCMARITTNCKANGCSWITTCKKNC 19	
RESULT	4	
LOCUS	A019667	396 bp DNA GSS 15-SEP-1998
DEFINITION	CIT-HSP-2381B2.TE CIT-HSP Homo sapiens genomic clone 2381B2,	
ACCESSION	A019667	genomic survey sequence.
NID	g3604029	
KEYWORDS	GSS.	
SOURCE	human.	
ORGANISM	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;	
	Primates; Catarrhini; Hominoidea; Homo.	

[illegible]

BASE COUNT	113 a	88 c	86 g	123 t
ORIGIN				
Query Match	55.5%	Score 111;	DB 15;	Length 410;
Best Local Similarity	50.0%;	Pred. No. 2,15e-04;		
Matches	19;	Conservative	7;	Mismatches 12; Indels 0; Gaps 0;
Db	144	GGCTTGACCGGGCAAGAGTCTGCTACCAAGCTGC	181	
Oy	7	GGNTTYACNGGNGCMGNAARMSNGCMGNARARYTNGC	44	
RESULT	6			
LOCUS	A1062871	520 bp	mRNA	EST
DEFINITION	GH02225.3:prime GH Drosophila melanogaster head P0T2 Drosophila			
ACCESSION	melanogaster cDNA clone GH02225 3prime, mRNA sequence.			
NID	A1062871			
KEYWORDS	G3338710			
SOURCE	ESM			
ORGANISM	fruit fly			
REFERENCE	Drosophila melanogaster			
AUTHORS	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
TITLE	Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;			
JOURNAL	Drosophilidae; Drosophila			
COMMENT	1 (bases 1 to 520)			
	Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,			
	Brockstein,P., Lewis,S. and Rubin,G.M.			
	BDGP/HMMI Drosophila EST Project			
	Unpublished (1997)			
	Contact: Harvey, D.			
	G. M. Rubin-Molecular and Cell Biology			
	University of California Berkeley			
	539 LSA, Berkeley, CA 94720-3200, USA			
	Fax: 510 643 9947			
	Email: http://fruitfly.berkeley.edu/EST , estfruitfly.berkeley.edu			
	Based upon the presence of a XhoI site followed by a run of 14 or			
	more T residues at the beginning of the sequence, this clone			
	probably contains an inverted insert. The sequence has been trimmed			
	and the T residues removed. hit genomic sequence DS02155			
	Plate: 22 row: C column: 1			
	High quality sequence stop: 408.			
FEATURES				
SOURCE				
	1..520			
	Location/Qualifiers			
	/organism="Drosophila melanogaster"			
	/note="Organ: head; Vector: P0T2; Site:1; EScore: Site:2;			
	XhoI: Sized fractionated cDNAs were directly ligated into			
	P0T2. Plasmid cDNA library."			
	/db_xref="taxon:7227"			
	/clone="GH02225"			
	/clone_1lb="GH Drosophila melanogaster head P0T2"			
	/sex="male and female"			
	/dev_strage="adulic"			
	/lab_host="DHS alpha"			
BASE COUNT	155 a	155 c	119 g	91 t
ORIGIN				
Query Match	55.5%	Score 111;	DB 15;	Length 520;
Best Local Similarity	41.7%;	Pred. No. 2,15e-04;		
Matches	20;	Conservative	10;	Indels 0; Gaps 0;

BASE COUNT	77 a	66 c	72 g	87 t
ORIGIN	/clone_11b="lambda uni-zap II"			
Query Match	53.0%	Score 106;	DB 20;	Length 302;
Best Local	Similarity 46.3%;	Prod. No. 2.50e-03;		
Matches	19;	Conservative 8;	Mismatches 14;	Indels 0; Gaps 0;
Db	127 TGGCTAGCTAGTTCCTGATCCATTTCCCTCCCTCTGAA	167		
Cp	50 TGRITNGCNARYTTNCKNGCNSWYTTNCKNGCNCGTAA	10		
RESULT	10			
LOCUS	B78873	358 bp	DNA	GSS
DEFINITION	C197978SK-81H4.TV C197978SK Homo sapiens genomic clone 81H4, genomic survey sequence.			
ACCESSION	B78873			
NID	g2865896			
KEYWORDS	GSS.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
AUTHORS	1 (bases 1 to 358)			
TITLE	Kim,U.-J., Adams,M.D. and Simon,M.I. Determination of clone end sequences of human Bacterial Artificial Chromosomes			
JOURNAL	Unpublished (1997)			
COMMENT	Other GSS: C197978SK-81H4.TP			
	Contact: Ung-Jin Kim			
	Caltech Genome Research Lab			
	California Institute of Technology			
	Division of Biology, MS 147-75, Pasadena, CA 91125, USA			
	Tel: 626 796 7066			
	Fax: 626 395 4901			
	Email: ung@ash.tree.caltech.edu			
	Clones are available from Research Genetics (info@resgen.com), BAC end search page:			
	http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html			
	Seq primer: 17			
	Class: BAC ends.			
FEATURES	Location/Qualifiers			
SOURCE	1..358			
	/organism="Homo sapiens"			
	/note="Vector: pBAC108L; Site_1: HindIII; Site_2: HindIII"			
	Caltech Human BAC Library A"			
	/db_xref="taxon:9606"			
	/clone="81H4"			
	/clone_11b="C197978SK"			
	/sex="Female"			
	/cell_type="Fibroblast"			
BASE COUNT	105 a	90 c	60 g	103 t
ORIGIN				
Query Match	53.0%	Score 106;	DB 26;	Length 358;
Best Local	Similarity 50.0%;	Prod. No. 2.50e-03;		
Matches	16;	Conservative 7;	Mismatches 9;	Indels 0; Gaps 0;
Db	200 GGGCTCTTTCTCGCTCCAGTAATCAACCA	231		
Cp	34 KNGCNSWYTTNCKNGCNCGTAAACNCNCR	3		
RESULT	11			
LOCUS	AO085207	376 bp	DNA	GSS
DEFINITION	HS-2270_A1_E08_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2270 Col=15 Row=1, genomic survey sequence.			
ACCESSION	AO085207			
NID	g34544424			
KEYWORDS	GSS.			
SOURCE	human.			

ORGANISM	Homo sapiens		
REFERENCE	Eukaryote; Metazoa; Chordata; Vertebrata; Mammalia; Euteria;		
ATTNORS	Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 376)		
JOURNAL	Mahairs,G.G., Zackrone,K.D., Smith,T., Tipton,S., Schmidt,S.,		
COMMENT	Tralcoff,R., Abajian,C., Blanchard,A., West,A. and Hood,L.E.		
	Construction of a Characterized Clone Resource for Genomic		
	Sequencing: Generation and Preliminary Analysis of 20,000 Sequence		
	Tagged Connectors		
	Unpublished (1997)		
	Contact: Mahairs GG, Wallace JC, Hood L		
	High Throughput Sequencing Center		
	University of Washington		
	401 Queen Anne Avenue North, Seattle, WA 98109, USA		
	Tel: (206) 616-3618		
	Fax: (206) 616-3887		
	Email: jwallace@u.washington.edu		
	Sequence Tagged Connector		
	Plate: 2270 row: I column: 15		
	Class: BAC ends		
	High quality sequence stop: 376.		
FEATURES	Location/Qualifiers		
SOURCE	1..376		
	/organism="Homo sapiens"		
	/note="Organ: Sperm; Vector: pBelBAC1; BAC Clones in		
	E-Coli DH10B"		
	/db_xref="taxon:9606"		
	/clone="Plate=2270 Col=15 Row=I"		
	/clone_lib="CIT Approved Human Genomic Sperm Library D"		
	/sex="male"		
BASE COUNT	126 a 100 c 75 g 75 t		
ORIGIN			
Query Match	53.0%; Score 106; DB 27; Length 376;		
Best Local	Similarity 39.2%; Pred 2.50e-03;		
Matches	20; Conservative 11; Mismatches 20; Indels 0; Gaps 0;		
Db	171 TTGGGGTGCTCACCCTGAGTCAAGAGTCGCTACACGCTGGCCACTAG 221		
QY	1 TTGGNGNTTYACGCGCMGNAARNSNCGMNAARYTGCNAACAR 51		
RESULT 12			
LOCUS	HUM073F02B 386 bp mRNA EST 21-MAY-1996		
DEFINITION	Human fetal brain cDNA 5'-end GEN-073F02, mRNA sequence.		
ACCESSION	D52294		
NID	9552350		
KEYWORDS	EST; EST(expressed sequence tag); Human fetal brain; similar to		
SOURCE	Known (May 29, 1995).		
ORGANISM	Homo sapiens cDNA to mRNA, clone_lib:Clontech human fetal brain		
	polyA+ mRNA (#6535).		
	Homo sapiens		
	Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata;		
	Vertebrata; Mammalia; Euteria; Primates; Catarrhini; Homnidae;		
	Homo.		
REFERENCE	1 (bases 1 to 386)		
ATTNORS	Fujiwara,T., Hirano,H., Katagiri,T., Kawai,A., Kuga,Y., Nagata,M.,		
	Okuno,S., Ozaki,K., Shimizu,F., Shinada,Y., Shinomiya,H.,		
	Takauchi,A., Takeda,S., Watanabe,T., Takahashi,E., Hirai,Y.,		
	Mekawa,H., Shin,S. and Nakamura,Y.		
TITLE	Unpublished (1995)		
JOURNAL	2 (bases 1 to 386)		
COMMENT	Fujiwara,T.		
	Direct Submission		
	Submitted (30-MAY-1995) to the DDBJ/EMBL/GenBank databases. Tsutomu		
	Fujiwara, Otsuka GEN Research Institute,Otsuka Pharmaceutical		
	Co.,Ltd; 463-10 Kagasuno Kawauchi-cho, Tokushima, Tokushima 771-01,		
	Japan (Tel:0886-65-2888, Fax:0886-37-1035)		
	Submitted (30-May-1995) to DDBJ by:		
	Tsutomu Fujiwara		
	Otsuka GEN Research Institute		

1. 403
/organism="Mus musculus"
/strain="C57Bl/6J"
/note="Organ: mammary gland; Vector: pW73D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5].
TGTTACCAATCTGAAGGAGGCGGGCGGAGTGTCTTTT
T 3'); double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia) digested with Not I and cloned into
the Not I and Eco RI sites of the modified pW73D vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M. Fatima
Donaldso."

```

/organism="Mus musculus"
/strain="C57BL/6J"
/notice=vector: pRT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo (dt) primer 15',
TGTTCACCATCTGCAAGTGGAGCGCGCCGCGGATGTTTTTTTTTTTTTTTTTTT
T 3'; double-stranded cDNA was ligated to Eco RI
adaptors (pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pRT73 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."
/db_xref="taxon:10090"
/clone="851105"
/clone_lib="Soares mouse mammary gland NBMHC"
/sex="male"
/dev_stage="4 weeks"
/lab_host="DH10B"
<1. .>428

```


BASE COUNT 139 a 83 c 84 g 122 t
 ORIGIN

Query Match 53.0%; Score 106; DB 7; Length 428;
 Best Local Similarity 48.5%; Pred. No. 2.50e-03;

Matches 16; Conservative 8; Mismatches 9; Indels 0; Gaps 0;

DB 35 GCAAGCACTGACCTGCAAGCTAGTAACACG 67
 19 GCNMGNAARNSGCMGNAARYTNCNAAACAR 51

RESULT 15

LOCUS FR0002492 437 bp DNA GSS 27-FEB-1997
 DEFINITION F.rubripes GSS sequence, clone 004C10BC10, genomic survey sequence.
 ACCESSION Z86275

NID 91883187

KEYWORDS GSS; genome survey sequence.

SOURCE Fugu rubripes.

ORGANISM

Fugu rubripes.

Eukaryota; Mitochondrial eukaryotes; Metazoa; Chordata;
 Vertebrata; Actinopterygii; Neopterygii; Teleostei; Euteleostei;
 Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae;
 Tetraodontidae; Fugu.

REFERENCE 1 (bases 1 to 437)
 Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umranta,Y.,
 Williams,G. and Brenner,S.

TITLE Direct Submission
 JOURNAL Submitted (18-FEB-1997) MRC Human Genome Mapping Project Resource
 Centre Hinxton, Cambridge, CB10 1SB. Email: blchelp@hgmrc.mrc.ac.uk

COMMENT Vector: pBluescript II KS
 V_type: phagemid
 PRIMER: KS

DESCR:

One pass dye-terminator sequencing of cosmid cloned genomic
 sequence.

FEATURES location/Qualifiers:
 source 1..437

/organism="Fugu rubripes"
 /db_xref="taxon:31033"

/clone_1lb="cosmid.004C10"
 /clone="004C10BC10"

BASE COUNT 114 a 114 c 94 g 108 t 7 others
 ORIGIN

Query Match 53.0%; Score 106; DB 26; Length 437;
 Best Local Similarity 51.1%; Pred. No. 2.50e-03;

Matches 24; Conservative 6; Mismatches 17; Indels 0; Gaps 0;

DB 17 CTGGTTNGCACTTCACTGTGTTTACCGCACCAGTGAGCCAC 63
 51 YTGRTTNGCNARYTNCNCGNSWYTTNCNCGNCNGTFAANCNC 5

Search completed: Tue Apr 27 10:16:42 1999
 Job time : 140 secs.

THIS PAGE BLANK (USPTO)

(TM)

Distribution rights by Oxford Molecular Ltd

in - protein database search, using Smith-Waterman algorithm

Fri Apr 16 14:21:22 1999; MasPar time 2.91 Seconds

t generated.

US-09-011-797-3

131 FEBRUARY 17

PAM 150

131922 seqs, 16180660 residues

Minimum	Match 08
Existing	first 45 summaries

1:geneSeq32
1:part1 2:part2 3:part3 4:part4 5:parts 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:parts 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29

mean 20.090; variance 71.751; scale 0.280

As the number of results predicted by chance to have a value greater than or equal to the score of the result being printed varied by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No
1	131	100.0	17	24	W25163	Human Opioid receptor	4.74e-06
2	60	45.8	700	23	W19992	Human GAP used to id	6.50e-01
3	59	45.0	321	20	W04560	Carnation ACC oxidase	8.00e-01
4	59	45.0	433	21	W14001	Enolase protein.	8.00e-01
5	58	44.3	39	28	W44735	Residues 317-375 of c	9.83e-01
6	58	44.3	270	28	W44734	Residues 317-375 of c	9.83e-01
7	58	44.3	319	23	W19537	Ethylene synthase fco	9.83e-01
8	57	43.5	618	25	W26541	Trypanosoma cruzi act	1.21e-02
9	57	43.5	345	12	R62504	Large polypeptide seq	1.21e-02
10	56	42.7	273	22	W20465	H. pylori outer membr	1.48e-02
11	56	42.7	273	22	W20465	H. pylori outer membr	1.48e-02
12	56	42.7	277	22	W20673	H. pylori outer membr	1.48e-02
13	56	42.7	343	22	W20333	H. pylori cytoplasmic	1.48e-02
14	56	42.7	343	22	W24634	H. pylori cytoplasmic	1.48e-02
15	55	42.0	16	17	R77675	Gluconylase N-termin	1.81e-02
16	55	42.0	318	22	W09880	ACC oxidase GAF-1.	1.81e-02
17	55	42.0	343	1	R02218	Sequence encoding Str	1.81e-02
18	55	42.0	466	19	R95688	Human GAP-SH3 domain	1.81e-02

19	55	42.0	62.1	17	R75674	Glucanase from <i>Arx</i>	1.8ie+02
20	55	42.0	82.1	22	W16611	<i>Candida albicans</i> topo	1.8ie+02
21	55	42.0	372.2	2	R10145	Cephalosporin antibio	1.8ie+02
22	54	41.2	243	17	R95271	Nisin nsi gene produ	2.22e+02
23	54	41.2	243	17	R95272	Nisin nsi gene produ	2.22e+02
24	54	41.2	1676	15	R77604	Pro-C5 polypeptide	2.22e+02
25	53	40.5	214	18	R95626	P. aeruginosa orf6-op	2.7ie+02
26	53	40.5	226	18	R95627	P. aeruginosa orf1-op	2.7ie+02
27	53	40.5	258	2	R10169	Vaccinia virus orf35	2.7ie+02
28	53	40.5	255	22	W17059	Vaccinia virus chemok	2.7ie+02
29	53	40.5	307	21	W13822	Yeast transcription r	2.7ie+02
30	53	40.5	320	18	R92096	Beta-ionone 4-methyl	2.7ie+02
31	53	40.5	350	1	P82053	Outer membrane protei	2.7ie+02
32	53	40.5	371	2	R30328	Xylose isomerase gene	2.7ie+02
33	53	40.5	387	1	R02217	Sequence encoding str	2.7ie+02
34	53	40.5	388	1	R02215	Sequence encoding str	2.7ie+02
35	53	40.5	388	1	P94687	DNA-deduced sequence	2.7ie+02
36	53	40.5	389	1	R02216	Sequence encoding str	2.7ie+02
37	53	40.5	394	1	P90423	<i>Streptomyces</i> spp. glu	2.7ie+02
38	53	40.5	722	18	R98327	Rat neuronal protein	2.7ie+02
39	53	40.5	1073	28	W37371	Human ST receptor pro	2.7ie+02
40	53	40.5	1073	28	W32063	Human ST receptor pro	2.7ie+02
41	53	40.5	1345	17	R96956	Multi-drug resistance	2.7ie+02
42	53	40.5	1417	17	R96955	Multi-drug resistance	2.7ie+02
43	53	40.5	1482	17	R96953	Multi-drug resistance	2.7ie+02
44	53	40.5	1489	17	R96952	Multi-drug resistance	2.7ie+02
45	53	40.5	1531	17	R93153	Multi-drug resistance	2.7ie+02

ALIGNMENTS

ID	RESULT	1	W25162 standard; peptide; 17 AA.
AC	W25163:		
DT	08-DEC-1997	(first entry)	
DE	Human opioid receptor-like 1 receptor ligand, nociceptin, fragment.		
KV	Opioid receptor-like 1; ORL1; ligand; pronociceptive; stress;		
KV	hyperalgesia; locomotor activity disease; anxiety; neuroendocrine;		
KW	memory; attention; sensory perception; learning; homeostasis;		
KW	hyperalgesia; nociceptin; endorphin; dynorphin A.		
OS	Synthetic.		
PN	W09707208-A1.		
PD	27-FEB-1997.		
PF	14-AUG-1996: BE0087.		
PR	15-AUG-1995: US-002368.		
PA	(UIBR.) UNIV LIBRE BRUXELLES.		
PI	Meunier J, Mollereau C, Parmentier M, Vassart G;		
DR	WPI: 97-165292/15.		
DR	N-PSDB; T79888.		
PT	Novel ligand for the opioid receptor-like receptor, nociceptin -		
PT	has pro-nociceptive properties, useful for treating or preventing		
PT	diseases related to e.g. stress, hyperalgesia, locomotor activity,		
PT	etc.		
PS	Claim 7; Page 27; 48pp; English.		
CC	W25162-W25164 are fragments of a ligand to human opioid receptor-like 1		
CC	(ORL1) receptor, designated nociceptin, which resembles the endorphin		
CC	dynorphin A. Inhibitors of nociceptin or its DNA, e.g. antisense		
CC	sequences and antibodies can be used as new types of drugs in the		
CC	control of various behaviours or functions. The inhibitors can be used		
CC	to prevent or treat a disease related to hyperalgesia, neuroendocrine		
CC	secretion, stress, locomotor activity, anxiety, instinctive behaviour.		
CC	learning and memory, homeostasis, hyperalgesia, hypoalgesia and/or		
CC	sensory perception.		
Sequence	17 AA:		
Query Match	100.0%;	Score 131;	DB 24; Length 17;
Best Local Similarity 100.0%;		Pred. No. 4.74e-06;	
Matches 17; Conservative		0; Mismatches 0;	Indels 0; Gaps 0
Db	1 fsefmrgyvlismgssq 17		
QY	1 fSEEMROTVLSMOSSQ 17		

CC 1s generated from a longer amino acid sequence (see W44734))
 SO Sequence 39 AA;

Query Match 44.3%; Score 56; DB 28; Length 39;
 Best Local Similarity 38.5%; Pred. No. 9.83e+01;
 Matches 5; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 11 fddfmrefrltel 23
 | : | | | : | : | :
 QY 1 FSEFMROYLVLSM 13

RESULT 6
 ID W44734 standard; peptide: 270 AA.
 AC W44734;

DT 18-MAY-1998 (first entry)
 DE Residues 307-576 of calpain I.
 KM Human; calpain I; antibody; detection; diagnosis; Alzheimer's disease.
 OS Homo sapiens.

FT Key Location/Qualifiers

FT Misc-difference 144

FT Misc-difference 261 /label= unknown

FT J0813596-A.

PD 07-MAY-1996

PF 16-AUG-1995; 208740.

PR 17-AUG-1994; US-292080.

PA (MCLE-) MCLEAN HOSPITAL CORP.

DR WPI: 98-021953/03.

PT Calpain-combining antibody - used for the diagnosis of Alzheimer's

PI disease

PS Disclosure: Fig 11: 11pp; Japanese.

CC This sequence represents amino acids 307-576 of the human calpain I.

CC The invention relates to an antibody which binds to amino acids 346-361

CC (W44733), corresponding to the domain III of the protein, and can be

CC used in the detection and diagnosis of Alzheimer's disease.

SO Sequence 270 AA;

Query Match 44.3%; Score 56; DB 28; Length 270;
 Best Local Similarity 38.5%; Pred. No. 9.83e+01;
 Matches 5; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 21 fddfmrefrltel 33
 | : | | | : | : | :
 QY 1 FSEFMROYLVLSM 13

RESULT 7
 ID W19537 standard; Protein; 319 AA.
 AC W19537;

DT 16-SEP-1997 (first entry)

DE Ethylene synthase from poplar plants induced by ozone.

KM Ozone tolerant; atmospheric pollution; photochemical oxidant;

OS Resistance.

PN J09096785-A.

PD 15-APR-1997

PF 06-OCT-1995; 284363.

PR 06-OCT-1995; JP-284363.

PA (TOYT) TOYOTA JIDOSHA KK.

DR WPI: 97-275444/25.

DR N-PSDB: T74024.

PT Ozone inducible ethylene synthase gene derived from poplar plants -

PT used for the production of ozone-tolerant plants

CC Claim 7: Page 6-7; 9pp; Japanese.

CC The present sequence represents a plant ethylene synthase, derived

CC from poplar plants, which is induced by ozone. The gene encoding

CC the ethylene synthase is used to develop a plant highly resistant

CC to ozone as a major photochemical oxidant in atmospheric pollution.

CC The plants use ozone for gene expression and in doing so reduce ozone

CC levels in lower atmosphere levels.

SO Sequence 319 AA;

Query Match 44.3%; Score 58; DB 23; Length 319;

Best Local Similarity 29.4%; Pred. No. 9.83e+01;

Matches 5; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Db 282 fddymklyaglkfagke 298
 | : | | | : | : | :
 QY 1 FSEFMROYLVLSMOSQ 17

RESULT 8
 ID W26541 standard; Protein; 618 AA.
 AC W26541;

DT 12-JAN-1998 (first entry)

DE Trypanosoma cruzi antigen.

KM Antigen; epitope; vaccine; protective immunity; Chagas disease;

OS Trypanosoma cruzi Tulihen strain C2.

PN N09718475-A1.

PD 22-MAY-1997

PF 14-NOV-1996; U18624.

PR 14-NOV-1995; US-557309.

PA (CORI-) CORIXA CORP.

PI Houghton RL, Iodes MJ, Reed SG, Skelky YAW.

DR WPI: 97-289413/26.

DR N-PSDB: T69167.

PT Diagnosing Trypanosoma cruzi infection by detecting antibodies to

PT novel antigens - which are useful in vaccines to provide protective

PT immunity against Chagas' disease

PS Disclosure: Page 88-91; 110pp; English.

CC This polypeptide sequence comprises a full-length antigen of

CC Trypanosoma cruzi, identified by sequencing a DNA clone (see T69167)

CC obtained by screening a Trypanosoma cruzi genomic expression library

CC with pools of sera from infected individuals. T. cruzi antigens

CC (see W26530-41), or epitope-containing repeat sequences (see W19094-

CC 102, W19079-86 and W26542-44) of native antigens, can be used in a

CC variety of immunassays for detecting T. cruzi infection in a

CC blood, serum, plasma, saliva, cerebrospinal fluid or urine sample.

CC The polypeptides are also useful in vaccines and pharmaceutical

CC compositions for inducing protective immunity against Chagas

CC disease. They can be produced by expression in transformed or

CC transfectected host cells.

SO Sequence 618 AA;

Query Match 43.5%; Score 57; DB 25; Length 618;
 Best Local Similarity 60.0%; Pred. No. 1.21e+02;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 189 ylvymqgave 198
 | | | | | : | : | :
 QY 8 YLVLSMOSQ 17

RESULT 9
 ID R62504 standard; Protein; 3457 AA.
 AC R62504;

DT 18-MAY-1995 (first entry)

DE Large polyprotein sequence of Maize Chlorotic Dwarf Virus.

KM MCDV; maize chlorotic dwarf virus; coat protein; plasmid vector;

KW plant resistance; viral infection resistance; maize plant;

OS Maize chlorotic dwarf virus; MCDV.

FT Key Location/Qualifiers

FT Region 439..829

FT cleavage_site 896..897

FT /label= "antigenic region of cp2"

FT /note= "may be used by animal picornavirus 3C

FT proteases"

FT peptide 897..1098

FT /note= "putative cp3 region, one of the structural

FT proteins"

FT peptide 897..911


```

FT      /note="N-terminus of cp3"
FT      1099..1113
FT      /note="N-terminus of cp1"
FT      1205..1483
FT      /note="antigenic region of cp1"
FT      1098..1099
FT      /label="di-peptide,cleavage_site"
FT      /note="may be used by animal picornavirus 3C
FT      proteases"
PN      WO9421796-A.
PD      29-SEP-1994.
PF      22-MAR-1994; US03028.
PR      24-MAR-1993; US-038768.
PA      (PION-) PIONEER HT-BRED INT INC.
PA      (USDA ) US SEC OF AGRIC.
PI      McMullen MD, Roth BA, Townsend R;
PI      WPI: 94-317016/39.
DR      N-PSDB: Q74694.
PT      DNA encoding maize chlorotic dwarf virus proteins - used to
PT      provide plants with resistance to the virus and related viral
PT      infections
PS      Disclosure: Page 18-32; 40pp; English.
CC      This sequence shows the large polyprotein of the Maize Chlorotic
CC      Dwarf Virus (MCDV). When cleaved it contains 3 structural proteins,
CC      cp1, cp2 and cp3. Depending on the exact location of cp2, the MCDV
CC      genome, can encode upto 78 kD of protein 5' of the capsid proteins
CC      (for which there are no corresponding animal picornavirus protein).
CC      The DNA is used for imparting resistance to MCDV or viruses to which
CC      MCDV infection or resistance provides cross-resistance, including
CC      maize dwarf mosaic virus strain A. Any or all of the three coat
CC      protein genes from MCDV can be used to provide protection for plants.
SQ      Sequence 3457 AA;

```

```

Query Match      43.5%; Score 57; DB 12; Length 3457;
Best Local Similarity 40.0%; Pred. No. 1.21e+02;
Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
Db      1945 fmeflkmyaalymdn 1959
Oy      1 FSEFMROYLVLSMOSS 15

```

```

RESULT 10
ID      W24665 standard; Protein; 273 AA.
AC      W24665;
DT      12-AUG-1997 (first entry)
DE      H. pylori outer membrane protein, 36126938.aa.
KM      Transmembrane; cytoplasmic; cell envelope; flagella; transport;
KM      secreted; periplasmic; chronic gastritis; duodenal ulcer disease;
KM      activator; inhibitor; bacterial life cycle; vaccine; immunise;
KM      detection; antisense; inhibition.
OS      Helicobacter pylori.
PN      WO9719098-A1.
PD      29-MAY-1997.
PF      15-NOV-1996; U18542.
PR      17-NOV-1995; US-561469.
PA      (ASTR ) ASTRA AB.
PI      Smith DH;
PI      WPI: 97-298052/27.
DR      N-PSDB: T77483.
PT      Helicobacter pylori nucleic acid sequences and related proteins -
PT      used for diagnostics and therapeutics
PS      Claim 18; Page 179-180; 235pp; English.
CC      This sequence represents an H. pylori outer membrane protein.
CC      Helicobacter pylori has been strongly linked to chronic gastritis and
CC      duodenal ulcer disease. The nucleic acid sequences of the invention
CC      are used to evaluate compounds, especially activators or inhibitors of
CC      bacterial life cycle, for the ability to bind an H. pylori nucleic acid
CC      sequence. The nucleic acid sequences, and corresponding proteins, are
CC      also useful for generating vaccines for immunising subjects against H.
CC      pylori or for use in detecting the presence of Helicobacter species in
CC      a sample. Antisense nucleic acid sequences of these sequences are
CC      used to inhibit expression of a gene from Helicobacter species. H.

```

```

CC      pylori whole genomic DNA was isolated and nebulised to a median size of
CC      2000 bp. Purified DNA fragments were blunt-ended and ligated to unique
CC      BstXI-linker adapters in 100-1000 fold molar excess. These linkers are
CC      complementary to the BstXI-cut PMP vectors, while the overhang is not
CC      self-complementary. Therefore the linkers will not concatemerise nor
CC      will the cut vector re-ligate itself easily. The linker-adaptor inserts
CC      were ligated to each of the 20 PMP vectors to construct a series of
CC      shotgun subclone libraries. The purified DNA samples were then
CC      sequenced.
CC      Note: The ORF/protein reference number for this sequence was obtained
CC      from the related specification, WO9640893.
SQ      Sequence 273 AA;

```

```

Query Match      42.7%; Score 56; DB 22; Length 273;
Best Local Similarity 35.7%; Pred. No. 1.48e+02;
Matches 5; Conservative 6; Mismatches 3; Indels 0; Gaps 0;
Db      2 efmkkfvalg11sa 15
Oy      3 EFMROYLVLSMOSS 16

```

```

RESULT 11
ID      W20465 standard; protein; 273 AA.
AC      W20465;
DT      29-JUL-1997 (first entry)
DE      H. pylori outer membrane protein, 36126938.aa.
KM      Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KM      binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KM      duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
OS      Helicobacter pylori.
PN      WO9640893-A1.
PD      19-DEC-1996.
PF      06-JUN-1996; U09122.
PR      07-JUN-1995; US-487032.
PR      01-APR-1996; US-630405.
PA      (ASTR ) ASTRA AB.
PI      Berglindh OT, Smith D, Mellgaard BJ;
PI      WPI: 97-052306/05.
DR      N-PSDB: T67803.
PT      Helicobacter pylori nucleic acid sequences and related
PT      polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT      infection, and to detect Helicobacter
PS      Claim 36; Page 635; 1481pp; English.
CC      The present sequence is a H. pylori outer membrane protein.
CC      The protein may be used in a vaccine to prevent or treat H. pylori
CC      infection or to identify H. pylori polypeptide binding compounds,
CC      useful as potential H. pylori life cycle activators or inhibitors.
CC      The genomic sequence of H. pylori (ATCC 53679) was determined from
CC      overlapping contigs generated by mechanically shearing the bacterial
CC      DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC      and the predicted coding regions defined by computer evaluation. To
CC      identify likely H. pylori antigens for vaccine development, the amino
CC      acid sequences predicted from various ORF were analysed for significant
CC      homology to other known or exported membrane proteins. Having identified
CC      and determined the sequences of interest, particular regions can be
CC      isolated from H. pylori by PCR amplification for recombinant polypeptide
CC      production, e.g. in E. coli hosts.
SQ      Sequence 273 AA;

```

```

Query Match      42.7%; Score 56; DB 22; Length 273;
Best Local Similarity 35.7%; Pred. No. 1.48e+02;
Matches 5; Conservative 6; Mismatches 3; Indels 0; Gaps 0;
Db      2 efmkkfvalg11sa 15
Oy      3 EFMROYLVLSMOSS 16

```

```

RESULT 12
ID      W20673 standard; protein; 277 AA.
AC      W20673;
DT      22-JUL-1997 (first entry)

```


DE H. pylori outer membrane protein, 04c11617orf2.
 KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
 KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
 OS Helicobacter pylori.
 PN W09640893-A1.
 PD 19-DEC-1996.
 PF 06-JUN-1996: U09122.
 PR 07-JUN-1995: US-487032.
 PR 01-APR-1996: US-630405.
 PA (ASTR) ASTRA AB.
 PI Berglindh OT, Smith D, Mellgaard BL;
 DR WPI: 97-052306/05.
 DR N-PSDB: T67926.
 PT Helicobacter pylori nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
 PT infection, and to detect Helicobacter
 PS Claim 56; Page 1094; 1481pp; English.
 CC The present sequence is a H. pylori outer membrane protein.
 CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds.
 CC Useful as potential H. pylori life cycle activators or inhibitors.
 CC The genomic sequence of H. pylori (ATCC 55679) was determined from
 CC overlapping contigs generated by mechanically shearing the bacterial
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
 CC and the predicted coding regions defined by computer evaluation. To
 CC identify likely H. pylori antigens for vaccine development, the amino
 CC acid sequences predicted from various ORF were analysed for significant
 CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide
 CC production, e.g. in E. coli hosts.
 SQ Sequence 277 AA;

Query Match 42.7%; Score 56; DB 22; Length 277;
 Best Local Similarity 35.7%; Pred. No. 1.48e+02;

Matches 5; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

Db 6 efmkrtvaqlgsa 19

QY 3 EFMROYLVLSMOSS 16

RESULT 13
 ID W20333 standard; protein: 343 AA.

AC W20333;

DE H. pylori cytoplasmic protein, 24816915.aa.

KW Vaccine; prevention; treatment; infection; identification;
 KW binding compound; bacterium; life cycle; activator; bacteria;
 KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;
 KW cytoplasmic.

OS Helicobacter pylori.

PN W09640893-A1.

PD 19-DEC-1996.

PF 06-JUN-1996: U09122.

PR 07-JUN-1995: US-487032.

PR 01-APR-1996: US-630405.

PA (ASTR) ASTRA AB.

PI Berglindh OT, Smith D, Mellgaard BL;

DR WPI: 97-052306/05.

DR N-PSDB: T67772.

PT Helicobacter pylori nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
 PT infection, and to detect Helicobacter

PS Claim 61; Pages 524-525; 1481pp; English.

CC The present sequence is a Helicobacter pylori cytoplasmic protein.
 CC The protein may be used in a vaccine to prevent or treat
 CC H. pylori infection or to identify H. pylori polypeptide binding
 CC compounds, useful as potential H. pylori life cycle activators or
 CC inhibitors. The genomic sequence of H. pylori (ATCC 55679) was
 CC determined from overlapping contigs generated by mechanically
 CC shearing the bacterial DNA. The sequences were analysed for ORF of

CC at least 180 nucleotides, and the predicted coding regions defined
 CC by computer evaluation. To identify likely H. pylori antigens for
 CC vaccine development, the amino acid sequences predicted from
 CC various ORF were analysed for significant homology to other known
 CC or exported membrane proteins. Having identified and determined
 CC the sequences of interest, particular regions can be isolated from
 CC H. pylori by PCR amplification for recombinant polypeptide
 CC production, e.g. in E. coli hosts.
 SQ Sequence 343 AA;

Query Match 42.7%; Score 56; DB 22; Length 343;
 Best Local Similarity 63.6%; Pred. No. 1.48e+02;

Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 71 fmeflkqrlvl 81

QY 1 FSEFMROYLVL 11

RESULT 14
 ID W24634 standard; protein: 343 AA.

AC W24634;

DE H. pylori cytoplasmic protein, 24816915.aa.

KW Chronic gastritis; duodenal ulcer disease; activator;
 KW inhibitor; bacterial life cycle; vaccine; immunisation; detection;
 KW antisense; inhibition; cytoplasmic.

OS Helicobacter pylori.

PN W09719098-A1.

PD 29-MAY-1997.

PF 15-NOV-1996: U18542.

PR 17-NOV-1995: US-561469.

PA (ASTR) ASTRA AB.

PI Smith DH;

DR WPI: 97-298052/27.

DR N-PSDB: T77452.

PT Helicobacter pylori nucleic acid sequences and related proteins -
 PT used for diagnostics and therapeutics

PS Claim 18; Pages 160-161; 235pp; English.

CC The present sequence is a Helicobacter pylori cytoplasmic
 CC protein.

CC H. pylori has been strongly linked to chronic gastritis and
 CC duodenal ulcer disease. The nucleic acid sequences of the invention
 CC are used to evaluate compounds, especially activators or inhibitors
 CC of bacterial life cycle, for the ability to bind an H. pylori
 CC nucleic acid sequence. The nucleic acid sequences, and
 CC corresponding proteins, are also useful for generating vaccines for
 CC immunising subjects against H. pylori or for use in detecting the
 CC presence of Helicobacter species in a sample. Antisense nucleic
 CC acid sequences of these sequences are used to inhibit expression of
 CC a gene from Helicobacter species. H. pylori whole genomic DNA was
 CC isolated and nebulised to a median size of 2000 bp. Purified DNA
 CC fragments were blunt-ended and ligated to unique BstXI-linker
 CC adapters in 100-1000 fold molar excess. These linkers are
 CC complementary to the BstXI-cut PMX vectors, while the overhang is
 CC not self-complementary. Therefore the linkers will not
 CC concatamerise nor will the cut vector re-ligate itself easily. The
 CC linker-adaptor inserts were ligated to each of the 20 PMX vectors
 CC to construct a series of shotgun subclone libraries. The purified
 CC DNA samples were then sequenced.

CC Note: The ORF/protein reference number for this sequence was
 CC obtained from the related specification, W09640893.

SQ Sequence 343 AA;

Query Match 42.7%; Score 56; DB 22; Length 343;
 Best Local Similarity 63.6%; Pred. No. 1.48e+02;

Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 71 fmeflkqrlvl 81

QY 1 FSEFMROYLVL 11


```

RESULT 15
ID R77675 standard; Peptide: 16 AA.
AC R77675;
DT 29-JUN-1996 (first entry)
DE Glucoamylase N-terminal secretion sequence.
KW glucoamylase; Yeast; Saccharomyces cerevisiae; production; starch;
   glucose; degradation; carbon source.
OS Arxula adeninivorans.
PN DE4423058-A.
PD 18-JAN-1996.
PE 15-JUL-1994; 425058.
PR 15-JUL-1994; DE-425058.
PA (PFLA-) INST PFLANZENGENETIK & KULTURPFLANZENFOR.
PI Bul M, Foerster S, Kunze G, Kunze I;
DR WPI; 96-069579/08.
PT Prodn. of heat stable glucoamylase in yeast able to utilise starch - by
   transformation with glucoamylase gene including secretion sequence from
   Arxula adeninivorans
PT Arxula adeninivorans
PS Claim 1; Page 7; 21pp; German.
CC The present sequence is the N-terminal sequence of a heat stable
   glucoamylase derived from Arxula adeninivorans. Yeast, e.g.
   Saccharomyces cerevisiae, can be transformed with the DNA (T08701)
   encoding the glucoamylase (R77674) for the production of the enzyme
   which is able to used starch as its carbon source. The enzyme is useful
   for degradation of starch, giving glucose as a final product. In
   yeast the DNA encoding the enzyme can be easily manipulated to
   produce the protein in high yield, which is otherwise difficult to
   do in A.adeninivorans.
CC do in A.adeninivorans.
SQ Sequence 16 AA;

Query Match 42.0%; Score 55; DB 17; Length 16;
Best Local Similarity 50.0%; Pred. No. 1.81e+02;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 1 mrqflalaaas 12
   |||:| |: :|
QY 5 MROYLVLMSQSS 16

```

Search completed: Fri Apr 16 14:21:38 1999
 Job time : 16 secs.

RC STRAIN-H37RV;
 RA PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
 RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-H37RV;
 RX MEDLINE: 96181548.
 RA PHILIPP W.J., POULET S., EIGLMEIER K., PASCOPELLA L.,
 RA BALSUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,
 RA COLE S.T.;
 RL PROC. NATL. ACAD. SCI. U.S.A. 93:3132-3137(1996).
 DR EMBL: 296800; E321660;
 DR PFAM: PF00440; tetr.
 SQ SEQUENCE 210 AA; 23396 MW; 2B866E2A CRC32;

Query Match 51.9%; Score 68; DB 2; Length 210;
 Best Local Similarity 43.8%; Pred. No. 1.10e+00;
 Matches 7; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

Db 106 PDELRFLYLSMERQ 121
 QY 2 SEFMROYLVLSMOSSQ 17

RESULT 3
 ID 051297 PRELIMINARY; PRT: 310 AA.
 AC 051297.
 DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
 DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
 DE CONSERVED HYPOTHETICAL INTEGRAL MEMBRANE PROTEIN.
 GN B80317.
 OS BORELIA BURGDORFERI (LYME DISEASE SPIROCHETE).
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; SPIROCHETES; SPIROCHAETALES;
 CC SPIROCHAETACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 35210 / B31;
 RX MEDLINE: 98065943.
 RA FRASER C.M., CASSENS S., HUANG W.M., SUTTON G.G., CLAYTON R.A.,
 RA LAHIGH R., WHITE O., KETCHUM K.A., DODSON R., HICKEY E.K., GWINN M.,
 RA DOUGHERTY B., TOMB O.-F., FLEISCHMANN R.D., RICHARDSON D.,
 RA PETERSON J., KERLAVAGE A.R., QUACKENBUSH J., SALBERG S., HANSON M.,
 RA VAN VUUT R., PALMER N., ADAMS M.D., GOCAYNE J.D., WEIDMAN J.,
 RA UTERBACK T., WATTHEY L., McDONALD L., ARTACH P., BOWMAN C.,
 RA GARLAND S., FUJII C., COTTON M.D., HORST K., ROBERTS K., HATCH B.,
 RA SMITH H.O., VENTER J.C.;
 RL NATURE 390:580-586(1997).
 DR EMBL: AE001138; G2688213; -.
 DR TIGR: B80317; -.
 SQ SEQUENCE 310 AA; 35680 MW; C8F8F32F CRC32;

Query Match 51.9%; Score 68; DB 2; Length 310;
 Best Local Similarity 50.0%; Pred. No. 1.10e+00;
 Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Db 228 FVFFRPYLVGLTSG 243
 QY 1 FSEFMROYLVLSMOSS 16

RESULT 4
 ID 064946 PRELIMINARY; PRT: 193 AA.
 AC 064946;
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
 DE ACC OXIDASE 2 (FRAGMENT).
 GN CS-ACO2.
 OS CUCUMIS SATIVUS (CUCUMBER).
 OC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE;
 CC VIOLELES; CUCURBITACEAE.
 RN [1]

RP SEQUENCE FROM N.A.
 RC TISSUE-FLORAL BUDS;
 RA RYAN J.E., ANDRALOJC P.J., WILLIS A.C., GUTTERIDGE S., PARRY M.A.J.;
 RL PLANT PHYSIOL. 116:1192-1192(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-FLORAL BUDS;
 RA PERL-TREVES R., KAHANA A., KORACH T., KESSLER N.;
 RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: AF033582; G3025695; -.
 FT NONTER
 SQ SEQUENCE 193 AA; 21553 MW; 1D8F81D1 CRC32;

Query Match 49.6%; Score 65; DB 10; Length 193;
 Best Local Similarity 29.4%; Pred. No. 3.45e+00;
 Matches 5; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

Db 155 FDDYMKLYMGIKFOAKE 171
 QY 1 FSEFMROYLVLSMOSSQ 17

RESULT 5
 ID 040063 PRELIMINARY; PRT: 339 AA.
 AC 040063;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
 DE HORDEUM VULGARE (BARLEY).
 OS EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; MONOCOTYLEDONEAE;
 CC CYPERALES; GRAMINEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-VAR. NK 1558;
 RA NAKANISHI H., SAKAGUCHI T., OKUMURA N., UMEHARA Y., NISHIZAWA N.,
 RA CHINO M., MORI S.;
 RL SUBMITTED (AUG-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: D37796; G520582; -.
 DR PFAM: PF00671; Fe_Asc_oxidored.
 SQ SEQUENCE 339 AA; 37732 MW; B38285E1 CRC32;

Query Match 49.6%; Score 65; DB 10; Length 339;
 Best Local Similarity 50.0%; Pred. No. 3.45e+00;
 Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Db 310 FRDEMRIVNVKLGS 325
 QY 1 FSEFMROYLVLSMOSS 16

RESULT 6
 ID 004076 PRELIMINARY; PRT: 310 AA.
 AC 004076;
 DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
 DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
 DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
 DE ACC-OXIDASE.
 GN AB-ACOL.
 OS VIGNA ANGIUARS (ADZUKI BEAN).
 OC EUKARYOTA; MITOCHONDRIAL EUKARYOTES; VIRIDIPLANTAE;
 CC CHAROPHYTA/EMBRYOPHYTA GROUP; EMBRYOPHYTA; MAGNOLIOPHYTA;
 CC MAGNOLIOPSIDA; RUTANAE; SAPINDALES; FABACEAE; PAPILIONOIDEAE;
 CC VIGNA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-EPICOTYLS.
 RA KANEITA T., KAKIMOTO T., SHIBAKA H.;
 RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: AB002667; D1020382; -.
 DR PFAM: PF00671; Fe_Asc_oxidored.
 SQ SEQUENCE 310 AA; 35334 MW; C9035224 CRC32;

Query Match 48.9%; Score 64; DB 10; Length 310;
Best Local Similarity 35.3%; Pred. No. 5.01e+00;
Matches 6; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

DB 282 FEDMYRLVATLKFOPE 298
1 FSEFMROYLVLSMOSSQ 17

RESULT 7
ID 058550 PRELIMINARY; PRT; 312 AA.
AC 058550:
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DE 312AA LONG HYPOTHETICAL PROTEIN.
GN PCH033.
OS PYROCOCCLUS HORIKOSHII.
OC ARCHAEA; EURYARCHAEOTA; THERMOCOCCALES; THERMOCOCCACEAE; PYROCOCCLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-073;
RA KAMARABAYASI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,
RA YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOGAWA A., NAGAI Y.,
RA SAKAI M., OGURA K., OTSUKA R., NAKAZAWA H., TAKAMITA M., OHEIKU Y.,
RA FUNAHASHI T., TANAKA T., KUDOH Y., YANAZAKI J., KUSHIDA N., OGUCHI A.,
RA AOKI K., YOSHIZAWA T., NAKAMURA Y., MASUCHI Y., SHIZUYA H.,
RA KIKUCHI H.;
RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AB009492; D1027933;
SQ SEQUENCE 312 AA; 35239 MW; A56CAE44 CRC32;

Query Match 48.9%; Score 64; DB 1; Length 312;
Best Local Similarity 69.2%; Pred. No. 5.01e+00;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

DB 141 FSGMROVLVLM 153
1 FSEFMROYLVLSM 13

RESULT 8
ID 040062 PRELIMINARY; PRT; 169 AA.
AC 040062:
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
GN IDS-3.
OS HORDEUM VULGARE (BARLEY).
OC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; MONOCOTYLEDONEAE;
OC CYPERALES; GRAMINEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-ROOT;
RA MORI S.;
RL SUBMITTED (APR-1992) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-ROOT;
RA NAKANISHI H., OKUMURA N., UMEHARA Y., NISHIZAWA N.K., CHINO M.,
RA MORI S.;
RL PLANT CELL PHYSIOL. 34:401-410(1994).
DR EMBL: D10058; G217877;
SQ SEQUENCE 169 AA; 18704 MW; 83B15F7B CRC32;

Query Match 48.1%; Score 63; DB 10; Length 169;
Best Local Similarity 43.8%; Pred. No. 7.24e+00;
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

DB 140 FRDEMRLVYVKGSA 155
1 FSEFMROYLVLSMOSS 16

RESULT 9
ID 061572 PRELIMINARY; PRT; 91 AA.
AC 061572:
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DE 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
GN HYPOTHETICAL 10.4 KD PROTEIN.
OS OSTERIAGIA OSTERIAGI.
OC EUKARYOTA; METAZOA; ACCELLOMATES; NEMATODA; SECERNENTEAE; RHABDITIDA.
RN [1]
RP SEQUENCE FROM N.A.
RA MOORE J., DEVANEY E.;
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AF052047; G2981080;
GN HYPOTHETICAL PROTEIN.
SQ SEQUENCE 91 AA; 10442 MW; 6448C6C7 CRC32;

Query Match 47.3%; Score 62; DB 5; Length 91;
Best Local Similarity 35.3%; Pred. No. 1.04e+01;
Matches 6; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

DB 44 YEFRRRYLVFEDSTE 60
1 FSEFMROYLVLSMOSSQ 17

RESULT 10
ID 024108 PRELIMINARY; PRT; 320 AA.
AC 024108:
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DE 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
GN 1-AMINOCYCLOPROPANE-1-CARBOXYLIC ACID OXIDASE.
RA NGACOL.
OS NICOTIANA GLUTINOSA (TOBACCO).
OC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE;
OC SOLANALES; SOLANACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-LEAF;
RA LEE S.H., KIM Y.S., LEE M.M., CHOI D.I., KIM W.T.;
RL SUBMITTED (APR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: U54565; G1314707;
GN PFAM: PF00671; FeAsc_oxidored.
SQ SEQUENCE 320 AA; 36288 MW; FDBDBE82 CRC32;

Query Match 47.3%; Score 62; DB 10; Length 320;
Best Local Similarity 29.4%; Pred. No. 1.04e+01;
Matches 5; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

DB 282 FDDYMKLVATLKFOPE 298
1 FSEFMROYLVLSMOSSQ 17

RESULT 11
ID 066456 PRELIMINARY; PRT; 406 AA.
AC 066456:
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DE 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
GN HYPOTHETICAL 47.6 KD PROTEIN.
RA AO_040.
OS AQUIFEX AEOLICUS.
OC AQUIFEXIA; AQUIFICALES; AQUIFICACEAE; AQUIFEX.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-YF5;
RA DECKERT G., WARREN P.V., GASTERLAND T., YOUNG W.G., LENOX A.L.,
RA GRAHAM D.E., OVERBEER R., SNEAD M.A., KELLER M., AUCAY M., HUBER R.,
RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;


```

RL  NATURE 392:353-358(1998).
RP  (2)
RP  SEQUENCE FROM N.A.
RC  STRAIN-VF5;
RA  DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
RA  GRAHAM D.E., OVERBECK R., SNEAD M.A., KELLER M., AUYAY M., HUBER R.,
RA  FELDMAN R.A., SHORT J.M., OLSON G.J., SMANSON R.V.;
RL  SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR  EMBL: A600670; G2982788; -.
KM  HYPOTHETICAL PROTEIN.
SQ  SEQUENCE 406 AA; 47559 MW; 3A69FCF8 CRC32;

Query Match
Best Local Similarity 47.3%; Score 62; DB 2; Length 406;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db  191 FEEFMRYLV 201
QY  1 FSEFMRYLV 11

RESULT 12
ID  Q54885 PRELIMINARY; PRT; 118 AA.
AC  Q54885;
DT  01-NOV-1996 (TREMBLREL. 01, CREATED)
DT  01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE  01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE  OREF10.
OS  STREPTOCOCCUS PNEUMONIAE.
OC  PROKARYOTA; BACTERIA; FIRMICUTES; GRAM-POSITIVE COCCI;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  TRANSPOSON-TRANSPOS TNS252;
RA  VIYAYAKUMAR M.N.;
RL  SUBMITTED (APR-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
DR  EMBL: L29324; G951053; -.
SQ  SEQUENCE 118 AA; 14251 MW; 1C07C914 CRC32;

Query Match
Best Local Similarity 46.6%; Score 61; DB 2; Length 118;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db  34 FSDFLRKSLLS 45
QY  1 FSEFMRYLV 12

RESULT 13
ID  P97544 PRELIMINARY; PRT; 312 AA.
AC  P97544;
DT  01-MAY-1997 (TREMBLREL. 03, CREATED)
DT  01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT  01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE  ER TRANSMEMBRANE PROTEIN.
GN  DRI 42.
OS  RATUS NORVEGICUS (RAT).
OC  EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC  EUTHERIA; RODENTIA.
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN-MISTAR; TISSUE=SMALL INTESTINE;
RX  MEDLINE: 97094703.
RA  BARIJA D., PLATEROTI M., NOBILI F., MUDA A.O., XIE Y., MORIMOTO T.,
RA  PEROZZI G.;
RL  J. BIOL. CHEM. 271:29928-29936(1996).
DR  EMBL: Y07783; E283078; -.
KM  TRANSMEMBRANE.
SQ  SEQUENCE 312 AA; 35318 MW; 5A8F58E2 CRC32;

Query Match
Best Local Similarity 46.6%; Score 61; DB 11; Length 312;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

```

```

Db  203 FSNFTMLVLYLQA 217
QY  1 FSEFMRYLVLSMOS 15

RESULT 14
ID  O64947 PRELIMINARY; PRT; 317 AA.
AC  O64947;
DT  01-AUG-1998 (TREMBLREL. 07, CREATED)
DT  01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DE  01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE  ACC OXIDASE 3.
GN  CS-AC03.
OS  CUCUMIS SATIVUS (CUCUMBER).
OC  EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE;
OC  VIOLALES; CUCURBITACEAE.
RN  [1]
RP  SEQUENCE FROM N.A.
RC  TISSUE=FLORAL BUDS;
RA  RYAN J.E., ANDRALOJC P.J., WILLIS A.C., GUTTERIDGE S., PARRY M.A.J.;
RL  PLANT PHYSIOL. 116:1192-1192(1998).
RN  [2]
RP  SEQUENCE FROM N.A.
RC  TISSUE=FLORAL BUDS;
RA  PERL-TREVES R., KAHANA A., KORACH T., KESSLER N.;
RL  SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR  EMBL: AF033583; G3025697; -.
SQ  SEQUENCE 317 AA; 35875 MW; 83DC530F CRC32;

Query Match
Best Local Similarity 46.6%; Score 61; DB 10; Length 317;
Matches 5; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

Db  283 FEDYMKLYGVKFOAKE 299
QY  1 FSEFMRYLVLSMOSQ 17

RESULT 15
ID  O49567 PRELIMINARY; PRT; 393 AA.
AC  O49567;
DT  01-JUN-1998 (TREMBLREL. 06, CREATED)
DT  01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT  01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE  HYPOTHETICAL 43.5 KD PROTEIN.
GN  F7J7.200.
OS  ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC  EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE;
OC  CAPRARALES; CROCIIFERAE.
RN  [1]
RP  SEQUENCE FROM N.A.
RA  BEVAN M., MURPHY G., DROST L., HALL C., HUDSON S., RIDLEY P.,
RA  BANCROFT I., MEWES H.W., MAYER K., SCHUELLER C.;
RL  SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
RN  [2]
RP  SEQUENCE FROM N.A.
RA  EU ARABIDOPSIS SEQUENCING PROJECT;
RL  SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR  EMBL: AL021960; E1253322; -.
KM  HYPOTHETICAL PROTEIN.
SQ  SEQUENCE 393 AA; 43457 MW; 3806ED3E CRC32;

Query Match
Best Local Similarity 46.6%; Score 61; DB 10; Length 393;
Matches 6; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db  311 OYLIGMONTE 321
QY  7 QYLVLSMOSQ 17

Search completed: Fri Apr 16 14:23:33 1999

```


Wed Apr 28 09:08:36 1999

US-09-011-797-3.rspt

Page 5

Job time : 35 secs.

THIS PAGE BLANK (USPTO)